

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2004, 09:56:03 ; Search time 6532 Seconds

(without alignments)

10866.785 Million cell updates/sec

Title:

Perfect score: 1501

Sequence: 1 gatggccactgctcctctg.....gtttccatcaagcttttccag 1501

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1501	100.0	7313	6	A69511
2	1501	100.0	7313	9	HSU65002
3	1499.4	99.9	1565	9	BC075047
4	1499.4	99.9	1565	9	BC075048
5	1265.8	84.3	1575	10	AF057366
6	1265.8	84.3	1575	10	AF057366
7	1260.4	84.0	142102	9	AC107952
8	1038	69.2	278377	2	AC129839
9	1031.6	68.7	182102	2	AC097274
10	1031.6	68.7	241148	10	AL807387
11	964.2	64.2	244955	2	AC123210
12	389.8	26.0	1820	9	BC023655
13	389.8	26.0	2335	6	CQ834146
14	389.8	26.0	2429	9	AK026936
15	389.8	26.0	4051	9	AF006005
16	389.8	26.0	5638	9	D83784
17	388.2	25.9	5267	10	AB051854
18	388.2	25.9	5474	10	AK129087
19	377.8	25.2	5643	6	CQ727766

20	361.8	24.1	1437	10	AF181262
21	356.2	23.7	161478	9	AC025538
22	351.6	23.4	1491	9	BC074814
23	351.6	23.4	2334	6	BD005470
24	351.6	23.4	2725	6	CQ716510
25	351.6	23.4	2828	6	CQ834144
26	351.6	23.4	2828	9	HS8635
27	351.6	23.4	3215	6	AX834095
28	351.6	23.4	3215	9	AK096331
29	351.6	23.4	3215	9	AK096331
30	350	23.3	2815	6	AX329760
31	350	23.3	3162	9	AK091707
32	350	23.3	4833	9	HSU72621
33	349.8	23.3	145414	9	HS8392M18
34	349.6	23.3	200329	10	AC078911
35	349.6	23.3	208614	10	AL807380
36	348	23.2	240510	2	AC134371
37	347	23.1	240336	2	AC094844
38	344.4	22.9	1874	5	AF186476
39	301.8	20.1	5028	10	RNU72620
40	295.8	19.7	2350	10	AF324471
41	294.4	19.6	2771	10	MMZINCPR
42	294.4	19.6	2790	6	BD005460
43	294.4	19.6	3732	10	MMZINCPR
44	294.4	19.6	3975	6	AX961848
45	294.4	19.6	3975	10	AF147785

ALIGNMENTS

RESULT 1	A69511	Sequence 1 from Patent WO9807748.	7313 bp	DNA	linear	PAT 07-MAY-1999
LOCUS	A69511					
DEFINITION	A69511					
ACCESSION	A69511					
VERSION	A69511.1	GI:4774166				
KEYWORDS						
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 7313)				
AUTHORS	PIAG GENE FAMILY AND TUMORIGENESIS					
TITLE	Patent: WO 9807748-A 1 26-FEB-1998;					
JOURNAL	VEN WILLEM JAN MARIE VAN DE (BE)					
FEATURES	Location/Qualifiers					
SOURCE	1..7313					
	/organism="Homo sapiens"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:9606"					
	481..1983					
	/note="unnamed protein product"					
	/codon_start=1					
	/protein_id="CAB42491.1"					
	/db_xref="GI:4774167"					
	/translation="MATVIPGLSEVDTQKVPSCRRKRGKTKPRKPNFPCQLCDKAFN					
	SVKELKLVSYSGHTGPRPYKCIQDCTKAPVKYKLRHMATHSPKTHKCNKCYKVFH					
	STGVLKHLHLTHDPNKETFKCECGKNTKLGFKRHLHAATSGDLTKVCLQTFE					
	RQVLEHLKHLHAGSGVKKHQCEHCDRFRYRKDVRRHVMVHTGRKDFLCQY					
	ACRFRKDLHTRHMKSHNQELLVKTEPFDLPFCNVSVPIKDELLPMSLPSE					
	LLSKPFTNLQNLNTPFQSMQSGSAHOMITTLPLGTCPIQMDVPHSHLSFKY					
	PFSTSTVAISIPKEQPLKGEISYLMELQGVSPSSQSSSSKSLKLPDQIGSLD					
	DGAGDLKSKSSISITDPLNTPALDPSOLFPLNGPPNPLSVSGLSGMSYSQBEAH					
	SSVQLPQTQDLDQDPANTIGLSLSAAPTSLSTSTTLPRFHAQF"					

Query Match 100.0%; Score 1501; DB 6; Length 7313;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGCCACTGCTCCTCGTGTGTTGTGTCAGAGTAGAGATACCCAGAAAGTCCTTC 60


```

/codon_start=1
/product="zinc finger protein PLG1"
/protein_id="AAC50995.1"
/db_xref="GI:1839160"
/translation="MAVVIPEGDLSEVRDQTKVPSGKKRGETKPRKNFPQQLCDKAFH
SVEKLKVSHTSGERPKYKICQDCTKAFYKSKYLQRHMAHPKETHKNCYCEKMFH
RKDHLKHLKTHDPNKFTKCEEGKNYNTKLGPKRHLAALHAATSGDLTKCVCLQTFE
STGVLLBHLKSHAGSKSGVKKEKHQCEHCGRDRPYTKDVRHMVHTGKDPCLQCYC
AQREFGRDHLTRMKHKNHQLKVKTEPDELDPDPTCNVSVPIKDELLPVMGSPSE
LLSKPFNTLQNLNPTFOSMSSGAHQMIITLPLMGTKCDIMDTVMHPSHHLSFKY
PFSSTVAISITPEKEQPLKEIETSLMELQGVFPSSQSSQASSKLGLDPQIGSLD
DGAGDLSSKGSISIDPLNTPALPDFSQLFNPLNGPPYNPPLSVGLSGLMGMSQEBAH
SSVSQDLPTQDQLQDNPANTTGLGSLHSLSAAFSSLSSTLTLPFRHQAFQ"
723..7313
/number=5

```

Query Match	100.0%;	Score 1501;	DB 9;	Length 7313;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1501;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GATGGCCACTGTCATTCTCTGGTGATTGTCAGAACTAAGAGATACCCAGAAAGTCCCTTC	60	
Db	480	GATGGCCACTGTCATTCTCTGGTGATTGTCAGAACTAAGAGATACCCAGAAAGTCCCTTC	539	
Qy	61	AGGGAACGCTAAGCGTGGTGAACCCAAACCAAGAAAACTTCTCTGCCACTGTGTGA	120	
Db	540	AGGGAACGCTAAGCGTGGTGAACCCAAACCAAGAAAACTTCTCTGCCACTGTGTGA	599	
Qy	121	CAAGGCCCTTAAACAGTGTGGAGAAATTAAAGTTCACTCTACTCTCACAGGAGAGAG	180	
Db	600	CAAGGCCCTTAAACAGTGTGGAGAAATTAAAGTTCACTCTACTCTCACAGGAGAGAG	659	
Qy	181	GCCTCAAGAGTGCCATACAACAAGACTGCACCAAGGCCCTTCTTCTAAGTACAAATTACA	240	
Db	660	GCCTCAAGAGTGCCATACAACAAGACTGCACCAAGGCCCTTCTTCTAAGTACAAATTACA	719	
Qy	241	AAGGCACATGGCTACTCATTTCTCTGAGAAAAACCCACAAGTGTAAATTATTTGTGAAAAAT	300	
Db	720	AAGGCACATGGCTACTCATTTCTCTGAGAAAAACCCACAAGTGTAAATTATTTGTGAAAAAT	779	
Qy	301	GTTTTACCGGAAAGATCATCTGAAGATACCTCCATACACAGACCCCTAACAAAGAC	360	
Db	780	GTTTTACCGGAAAGATCATCTGAAGATACCTCCATACACAGACCCCTAACAAAGAC	839	
Qy	361	GTTTAAAGTGCGAAGAAATGTGCGAAGAACTACAATACCAAGCTTGGATTAAACCTCACTT	420	
Db	840	GTTTAAAGTGCGAAGAAATGTGCGAAGAACTACAATACCAAGCTTGGATTAAACCTCACTT	899	
Qy	421	GGCCTTGCATGCCGCAACAGTGGTGAACCTTAAATCTCATGCGGCAAGTCGTCGTGGGGT	480	
Db	900	GGCCTTGCATGCCGCAACAGTGGTGAACCTTAAATCTCATGCGGCAAGTCGTCGTGGGGT	959	
Qy	481	AAGCAGCGGAGTGCTTCTGGAGCACCTTAAATCTCATGCGGCAAGTCGTCGTGGGGT	540	
Db	960	AAGCAGCGGAGTGCTTCTGGAGCACCTTAAATCTCATGCGGCAAGTCGTCGTGGGGT	1019	
Qy	541	TAAAGAAAAAAGACACCACTGCGAAACATTTGTGATCGCCGGTTTACACCCGAAGGATGT	600	
Db	1020	TAAAGAAAAAAGACACCACTGCGAAACATTTGTGATCGCCGGTTTACACCCGAAGGATGT	1079	
Qy	601	CCGAGACACATGGTGGTGACACCTGGAAGAAAGGACTTCTCTGTGAGTATTGTGCACA	660	
Db	1080	CCGAGACACATGGTGGTGACACCTGGAAGAAAGGACTTCTCTGTGAGTATTGTGCACA	1139	
Qy	661	GAGATTTCGGCGAAAGGATCACTGTGACTCGACATATGAAGAGAGTCAAAATCAAGAGCT	720	
Db	1140	GAGATTTCGGCGAAAGGATCACTGTGACTCGACATATGAAGAGAGTCAAAATCAAGAGCT	1199	
Qy	721	TCTGAAGGTCAAAAAGAACCACTGATTTCCTTGTGACCCATTACTGTCAATGTGTCTGT	780	
Db	1200	TCTGAAGGTCAAAAAGAACCACTGATTTCCTTGTGACCCATTACTGTCAATGTGTCTGT	1259	

RESULT		3		
LOCUS	BC075047			
DEFINITION	Homo sapiens pleiomorphic adenoma gene 1, MGC:103361 IMAGE:30915361), complete cds.	linear mRNA PRI 04-AUG-2004		
ACCESSION	BC075047	1565 bp		
VERSION	BC075047.2	GI:50960384		
KEYWORDS	MGC.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1565) Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Stausen,B.D., Collins,F.S., Wagner,L., Shennen,C.W., Schuler,G.D., Hoptkins,R.F., Zeeberg,B., Buetow,K.H., Schaefet,C.F., Bhat,N.K., Diatchenko,L., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,M., Soares,M.B., Farmer,A.A., Rubin,G.M., Hong,L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Casavant,T.L., Toshiyuki,S.,			
AUTHORS				

LLSKPFTNTLQNLVNTPQSMQSSGAHQMTITLPLGMCPTIDMTVTHPSHHLSFKY
 PFSTSYATSIPKEQPLKGEISYLEMELQGVSPSSQSDQASCKLGDLPQIGSLD
 DGAGLSKXSISIDPNTPALDLSQLEFNIPPLPPYPLSVGSLGMSYQEEAH
 QSVQLOPPOTDLODPANTIGLSLHSLSAETSSUSTSITLTPRFHQAFQ"

ORIGIN

Query Match	99.9%	Score 1499.4;	DB 9;	Length 1565;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1500;	Conservative	0;	Mismatches	1;
			Indels	0;
			Gaps	0;

QY
QY

DB 42 GATGGCCACAGTGCATTCCTCGGGAATGATCACTGATTTGTTTGTCTTCTTGCCAACTGTGTGA 120

61 AGGGAAACGTAAAGCGTGGTGAACAACCAACCAAGAAAAAATTTTCTTCTTGCCAACTGTGTGA 120

QY

DB	102	AGGGAACGTAAGCTGCTGAAACCAACCAAGAAAAAACTTTCTCTTGCAACTGTGTGA	16
OV	121	CAAGGCCCTTTAACAGTGTGTGAGAAATTTAAAGGTTCACTCTCTCTACACAGAGAGAG	180

Db
162 CAAGGCCCTTTACAGTGTGGAGAAATTAAGGTTCACTCCTACTCTTCACACAGGACGAG 221

[illegible]

QY	241	AAGGCACATGGCTACTCAATTCCTCTGAGAAAACCCACACAGTGTAAATATTTGTGAGAAAAAT	300
DB	282	AAGGCACATGGCTACTCAATTCCTCTGAGAAAACCCACACAGTGTAAATATTTGTGAGAAAAAT	341

301 GTTTTACCGGAAAGATCATCTGAGGAATCACCTCCATACACGACCCCTAACAAAGAGAC 360
QY
401

	GTTTACCGGAGAGATCATCTGGAGATTCTTC	342
DB		
	GTTTAAAGTGCGAAGAATGTGCCAAGCAACTCAATCCAGCTTGATTTAAACGTCACCTT	361
QY		

Db	402	GTTHAGTGCAGAAATGTCGCAAGAACTACAATACCARGCTTGGANTTAAACGTCACCTT	46
	421	GGCCTTGCATGCCGCAACAAGTGGTGACCTCACCTGTAAAGTATGTTTGGCAAACTTTTGA	480
Qv	421	GGCCTTGCATGCCGCAACAAGTGGTGACCTCACCTGTAAAGTATGTTTGGCAAACTTTTGA	480

db

462 GGCCCTGCAATGCCCGAACAACTGGTGACCTCACCTGTAGGTATGTTTGCACAACCTTTGA 521

Db

462 GGCCCTGCAATGCCCGAACAACTGGTGACCTCACCTGTAGGTATGTTTGCACAACCTTTGA 540

QY
481 AAGCACGGGAGCGCTTCACGATCCTTCTCATGCAGCAAGTCGTCTGTTGGGT 581

Dd
522 AAGCACGGGAGTGCTTCTGGAGCACCTTAATCTCATGCAGCAAGTCGTCTGTTGGGT 581

Qy	541	TAAAGAAAAAAGCACCAGTCGGACAATTGTCATCGCGGTTCTACACCCGAAAGGATG	590
Db <th>582</th> <th>TAAAGAAAAAAGCACCAGTCGGACAATTGTCATCGCGGTTCTACACCCGAAAGGATG</th> <th>641</th>	582	TAAAGAAAAAAGCACCAGTCGGACAATTGTCATCGCGGTTCTACACCCGAAAGGATG	641

Qy
601 CCGGAGACACATGGTGGTGCACACTCGAAGAAAGGACTTCCTCTGTCTAGTATTGTGCGACA
650

DB	Qy	72
642	CGGAGACACATGGGAGGAGCAGACAGTCAATCAAGAGCT	
661	GAGATTTGGGCGAAAGGATCACTGATCGACATATGAAGACAGTCAATCAAGAGCT	

702	GAGATTTGGCGGAAAGGATCACCTGACTCGACATATGAAGAAGAGTCAAAACAAGAGCT	78
Dδ		
721	TCTGAAGGTCAAAACAGAAACGAGTGGATTTCTTTCACCATATTCCTGCAATGTGTCTCT	79
Oγ		

762 TCTGAAGGTCAAACAGAAACCAAGTGGATTTCCTTGACCATTTACCTGCAATGTCGT 82

Ddb

QY GGCATATAAAGACGAGCCCTTTCCTTTGGTGTCTTCCCTCAGTGAACTGTTATCAA 88
781 |||||
Db GCCTATAAAGACGAGCTCTTCGCGGTGAATGCTTACCCTTCAGTGAACTGTTATCAA 89

Accession	Gene	Accession	Gene
QY	841	GCATTTACAAACACTTTGCAGTTAAACCTCTACAACACTCCATTTCCATCCATGCAGAG	94
Db	882	GCCATTTCAAACACTTTTCGAGTTAAACCTCTACAACACTCCATTTTCAGTCCATGCAGAG	94

[illegible]

DB
942 CCGGAGTCGCCACCAATGCTGGTATTGTTT

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Faney, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzyzinski, M.I., Skalski, U., Smalius, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1565)
Director WGC Project.
Direct Submission
Submitted (25-JUN-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 4, 2004 this sequence version replaced gi:49902411.
Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Genome Sequence Centre, British Columbia Cancer
Centre
cDNA Library Preparation: British Columbia Cancer Research Center
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
cDNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabh, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smailus, Jeff Scott, Miranda Tsai, George Yang, Jackie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: TPRU Plate: 3 Row: A Column: 2.

```

Location/Qualifiers
1. .1565
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="MGC:103961 IMAGE:30915361"
/tissue_type="Brain, PCR rescued clones"
/clone_lib="NIH MGC_264"
/lab_host="DH10B"
/note="vector: pCR4 Topo TA"
1. .1565
/gene="PLAG1"
/note="synonyms: SGPA, PSA"
/db_xref="LocusID:5324"
/db_xref="MIM:603026"
43. _1545

```

```

/ gene="PLAG1"
/ codon_start=1
/ product="pleomorphic adenoma gene 1"
/ protein_id="AAH75047.1"
/ db_xref="GI:49902412"
/ db_xref="LOCUSID:5324"
/ db_xref="MIM:603026"
/ translation="WATVPGDLSVDRDQKVPSSGRKRGTGTPKRNFPQQLCDKAFN
SYEKLUKVSYSHTGPEPYKCIQQDCTAFYSKYKLQRHMATSPKTHKNCYKRMFH
RKDHLKHLNHTDHPKE"FKCECGKNYTKLGPKHLAAHATSDGLTCVCLQTFE
STGVLLHLKSHAGSSGVKEKHQCEHCDRRYRTRKDVRRHNVHTGKDKLCOYC
AQRFGRKDLHTRMKKSHNQBLKVKTEPVDLPDFFTCNLSVPIKLELLPWLPSSE

```

QY 961 TATGGACACTGTTTCATCCCTCTCACACCTTTCTTCAATATACCGTTCAGTTCACCTC 1020
Db |||||||
QY 1002 TATGGACACTGTTTCATCCCTCTCACACCTTTCTTCAATATACCGTTCAGTTCACCTC 1061
Db |||||||
QY 1021 ATATGCAATTTCTATTCCTGAAAGAAAGAACAGCCATTAAAGGGGGAATTCAGAGTTACCT 1080
Db |||||||
QY 1062 ATATGCAATTTCTATTCCTGAAAGAAAGAACAGCCATTAAAGGGGGAATTCAGAGTTACCT 1121
Db |||||||
QY 1081 GATGGAGTTACAGGTGGCGTGCCTCTTCATCCCAAGATTCCTCAAGCATGTCATCATC 1140
Db |||||||
QY 1122 GATGGAGTTACAGGTGGCGTGCCTCTTCATCCCAAGATTCCTCAAGCATGTCATCATC 1181
Db |||||||
QY 1141 TAAGCTAGGTTGGATTCCTCAGATTGGTCCCTAGATGATGTCAGGAGACCTCTCCCT 1200
Db |||||||
QY 1182 TAAGCTAGGTTGGATTCCTCAGATTGGTCCCTAGATGATGTCAGGAGACCTCTCCCT 1241
Db |||||||
QY 1201 ATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAACACACACGAGCTTCGATTTCTCA 1260
Db |||||||
QY 1242 ATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAACACACACGAGCTTCGATTTCTCA 1301
Db |||||||
QY 1261 GTTGTGTTAATTCATACCTTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCCT 1320
Db |||||||
QY 1302 GTTGTGTTAATTCATACCTTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCCT 1361
Db |||||||
QY 1321 TGGATGAGCTATTCAGGAGAGACCATTTCTGTTCCTCCAGTCCCCACACAAAC 1380
Db |||||||
QY 1362 TGGATGAGCTATTCAGGAGAGACCATTTCTGTTCCTCCAGTCCCCACACAAAC 1421
Db |||||||
QY 1381 ACAGGATCTTCAGGATCTCCGAAACACTATAGGCTTGCTGCTCTGCACTCACTGTCAGC 1440
Db |||||||
QY 1422 ACAGGATCTTCAGGATCTCCGAAACACTATAGGCTTGCTGCTCTGCACTCACTGTCAGC 1481
Db |||||||
QY 1441 AGCTTTCACAGCTGTTTAAAGCAAGTACCACCTCCAGCTTCCATCAAGCTTTTCA 1500
Db |||||||
QY 1482 AGCTTTCACAGCTGTTTAAAGCAAGTACCACCTCCAGCTTCCATCAAGCTTTTCA 1541
Db |||||||
QY 1501 G 1501
Db 1542 G 1542

BC075048 1565 bp mRNA linear PRI 04-AUG-2004
Homo sapiens pleiomorphic adenoma gene 1, mRNA (cdna clone
MGC:104123 IMAGE:30915577), complete cds.
BC075048
BC075048.2 GI:50960210
MGC.
Homo sapiens (human)
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1565)
Klausner, R.D., Collins, F.S., Wagner, L., Hense, L.H., Derge, J.G.,
Strausberg, R.L., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Altschul, S.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Hopkins, R.F., Mariani, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Diatchenko, L., Mariani, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abrahamson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Schein, A.,
Schnurch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

REMARK
COMMENT

FEATURES
source

gene

CDS

ORIGIN

Query Match
Best Local
Matches 1500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1565)
Director MGC Project.
Direct Submission
Submitted (25-JUN-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 4, 2004 this sequence version replaced gi:49902222.
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Genome Sequence Centre, British Columbia Cancer
Center
cDNA Library Preparation: British Columbia Cancer Research Center
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,
Parvaneh Saeedi, JR Santos, Angeliq Schnerch, Ursula Skalska,
Duane Smalley, Jeff Stott, Miranda Tsai, George Yang, Jacques
Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRBU Plate: 5 Row: C Column: 2.
Location/Qualifiers

1..1565
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:104123 IMAGE:30915577"
/tissue_type="Brain, PCR rescued clones"
/clone_lib="NIH_MGC_264"
/lab_host="DH10B"
/notes="Vector: pCR4 Topo TA"
1..1565
/gene="PLAG1"
/note="synonyms: SGPA, PSA"
/db_xref="LocusID:5324"
/db_xref="MIM:603026"
43..1545
/genes="PLAG1"
/codon_start=1
/product="pleiomorphic adenoma gene 1"
/protein_id="AAH75048.1"
/db_xref="GI:49902223"
/db_xref="LocusID:5324"
/translation="WATVPGDLSEVDTQKVPKGRKRGKTKPRKNFPCQLCDKAFN
SVEKLKVSYSHTGEPYKCIQDCTKAFVSKYKLQRHATHSPKTHKNCYCEKMFH
RDKLKNLHLDHPKERTKCECKNNTKLGRKRLHAAHATSGDLTKCVLCLOTFE
SGVLKHLKSHAGSKSGGVKQCECHCDRFRYTRKDVRRHVVHVRGKDFLCQYC
AQGRKDLHTRHMKSHNQLLVKVPVDFDPDFTCNVSVPIKDELLPWSLPSSE
LLSFPTNLQNLNTPFQSMQSSGAHOMITLPLGTCPIQDMDVHPHSLFSKY
PFSSTSYAISPEKEQPLKGEISYLMELQGVPSQSSQSSSKGLGDPQIGSLD
DGAGLSLSSKSSISIDPLNTPALDFSQLFNPPLNPLSVSLGMSYSQBEAH
SSVSQLPQTDQDLPANTIGLSLHLSAAFTSLSTSTLTPRHQAQ"

Query Match
Best Local
Matches 1500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1122 GATGGAGTTACAAAGTTGGCGTGCCTCTTTCATCCCAAGATTCTCAAGCATCGTCATCATC 1181
1141 TAAGCTAGGGTTGGATCCCTCAGATTGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCT 1200
1182 TAAGCTAGGGTTGGATCCCTCAGATTGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCT 1241
1201 ATCCAAAAGCTCTATCTCCATCAGTGAACCCCTTAAACACACACACAGCATTTGATTTCTCA 1260
1242 ATCCAAAAGCTCTATCTCCATCAGTGAACCCCTTAAACACACACACAGCATTTGATTTCTCA 1301
1261 GTTCTTTAAATTTTCAATCTTTAAATGGTCTCCCTTAAATCTCTATCAGTGGGAGCCT 1320
1302 GTTCTTTAAATTTTCAATCTTTAAATGGTCTCCCTTAAATCTCTATCAGTGGGAGCCT 1361
1321 TGAATGAGCTATTTCCAGGAAGACACATTTCTCTTTCCAGCTCCCCACACAAAC 1380
1362 TGAATGAGCTATTTCCAGGAAGACACATTTCTCTTTCCAGCTCCCCACACAAAC 1421
1381 ACAGATCTTCAGGATCTCGAAACACATATAGGGCTTGGGTCTCTGACTCATCTGACG 1440
1422 ACAGATCTTCAGGATCTCGAAACACATATAGGGCTTGGGTCTCTGACTCATCTGACG 1481
1441 AGCTTTCCAGGAGCTTTAAGCACAGTACACCTCCACAGTTTCCATCAAGCTTTTCA 1500
1482 AGCTTTCCAGGAGCTTTAAGCACAGTACACCTCCACAGTTTCCATCAAGCTTTTCA 1541
1501 G 1501
1542 G 1542

RESULT 5
AF057366 1575 bp mRNA linear ROD 01-JAN-2000
LOCUS Mus musculus C2H2 zinc finger protein FLAG1 (Flag1) mRNA, complete cds.
DEFINITION
ACCESSION AF057366
VERSION AF057366.1 GI:6650194
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1575)
AUTHORS Kas,K., Voz,M.L., Roijer,E., Astrom,A.K., Meyen,E., Stenman,G. and Van de Ven,W.J.
TITLE Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations
JOURNAL Nat. Genet. 15 (2), 170-174 (1997)
MEDLINE 97172974
PUBMED 9020842
REFERENCE 2 (bases 1 to 1575)
AUTHORS Hensen,K., Voz,M.L., Van de Ven,W.J.M. and Kas,K.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1998) Laboratory of Molecular Oncology, Center of Human Genetics, Herestraat 49, Leuven 3000, Belgium
FEATURES
source
1..1575
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129/SV"
/db_xref="taxon:10090"
/map="t(3;8)(p21;q12)"
/tissue_type="pleiomorphic adenoma"
1..1575
/gene="Flag1"
/note="synonym: pleiomorphic adenoma gene 1"
21..1520
/gene="Flag1"
/codon_start=1
/product="C2H2 zinc finger protein FLAG1"
/protein_id="AAF21762.1"

42 GATGGCACTGTCTATCTCTGGTGTATTTGTCAGAGTAAGAGATACCCAGAAAGTCCCTTC 101
61 AGGGAACGTAAGCGTGGTGAACCAACCAAGAAAAATTTTCCTTGGCAACTGTGTGA 120
102 AGGGAACGTAAGCGTGGTGAACCAACCAAGAAAAATTTTCCTTGGCAACTGTGTGA 161
121 CAAGGCCCTTTAACAAGTGTGAGAAATTAAGGTTTCACTCTTACCTCTCACAGAGAGAG 180
162 CAAGGCCCTTTAACAAGTGTGAGAAATTAAGGTTTCACTCTTACCTCTCACAGAGAGAG 221
181 GCCCTACAAGTGCATACAAAGACTGCACCAAGCCCTTTGTTTCTTAAGTACAAATTACA 240
222 GCCCTACAAGTGCATACAAAGACTGCACCAAGCCCTTTGTTTCTTAAGTACAAATTACA 281
241 AAGGCACATGCTACTCATTTCTCTGAGAAACCCCAAGTGTATTAATTTGTGAGAAAT 300
282 AAGGCACATGCTACTCATTTCTCTGAGAAACCCCAAGTGTATTAATTTGTGAGAAAT 341
301 GTTTCACCGGAAAGATCTCTGAAGATCACTCTCATACACAGCACCCCTTAAACAAAGAGAC 360
342 GTTTCACCGGAAAGATCTCTGAAGATCACTCTCATACACAGCACCCCTTAAACAAAGAGAC 401
361 GTTTCACCGGAAAGATCTCTGAAGATCACTCTCATACACAGCACCCCTTAAACAAAGAGAC 420
402 GTTTCACCGGAAAGATCTCTGAAGATCACTCTCATACACAGCACCCCTTAAACAAAGAGAC 461
421 GGCCTTGATGCGGCAACCAAGTGGTGAACCTCTGAGGATGATTTGTTTCAAACTTTGA 480
462 GGCCTTGATGCGGCAACCAAGTGGTGAACCTCTGAGGATGATTTGTTTCAAACTTTGA 521
481 AAGCACGGAGTCTCTGGAGCACCTTAAATCTCATGAGCAAGTCTGTTGGGGT 540
522 AAGCACGGAGTCTCTGGAGCACCTTAAATCTCATGAGCAAGTCTGTTGGGGT 581
541 TAAAGAAAAAAGACACAGTGGCAACATTGTGATCGCCGGTTCTACACCCGAAAGATGT 600
582 TAAAGAAAAAAGACACAGTGGCAACATTGTGATCGCCGGTTCTACACCCGAAAGATGT 641
601 CCGGAGACACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTGAGTATTTGTCACA 660
642 CCGGAGACACATGTTGGTGCACACTGGAAGAAAGGACTTCTCTGTGAGTATTTGTCACA 701
661 GAGATTGGCGGAAAGATCACTGACTCGACATATGAAGAAAGTCTCAATCAAGAGCT 720
702 GAGATTGGCGGAAAGATCACTGACTCGACATATGAAGAAAGTCTCAATCAAGAGCT 761
721 TCTGAGGTCAAAACAGAACAGTGGATTTCTTTGACCCCAATTTACCTGCAATGTCTGT 780
762 TCTGAGGTCAAAACAGAACAGTGGATTTCTTTGACCCCAATTTACCTGCAATGTCTGT 821
781 GCCTATAAAGACGAGCTCTCTCCGGTGAATGCTTACCTTCCAGTGAAGTATCAAA 840
822 GCCTATAAAGACGAGCTCTCTCCGGTGAATGCTTACCTTCCAGTGAAGTATCAAA 881
841 GCCTATAAAGACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 900
882 GCCTATAAAGACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 941
901 CTCGGGATCTGCCACCAAAATGATCAACACTTTACCTTTGGGAATGATGCCCAATAGA 960
942 CTCGGGATCTGCCACCAAAATGATCAACACTTTACCTTTGGGAATGATGCCCAATAGA 1001
961 TATGACACTGTTATCCCTCTCAACACTTTTCAATATCGTTTCACTTACCTTC 1020
1002 TATGACACTGTTATCCCTCTCAACACTTTTCAATATCGTTTCACTTACCTTC 1061
1021 ATATGCAATTTCTATCTCTGAAAGAAAGACAGCAATTAAGGGGAAATTTGAGATTTACCT 1080
1062 ATATGCAATTTCTATCTCTGAAAGAAAGACAGCAATTAAGGGGAAATTTGAGATTTACCT 1121
1081 GATGGAGTTACAAAGTGGCGTGCCTCTTTCATCCCAAGATTCTCAAGCATCGTCATCATC 1140

/db_xref="GI:6650195"		/translation="MATVPGDLSFVDTQKAPSGKRRKSGSPKRNPFQCLCDKAFN	
SVKELVHFSHTGSRPKYKCTHODCTKAFVSKYKQORHATHSPETHKNCERKMFH		860 GCCATTCAAAACACATTTGCGATTAAACCTCTATAACATCCATTCAGTCCATGCAGAG	
RDHLKHLUHTDHPNKTFCBCEGKSYNTKLFGRHLALHAATSDGLTKVCLONF		901 CTGGGATCTGCCACCAAAATGATCAAACTTTACCTTTGGGAAATGACATGCCCAATAGA	
STGVLEHLHSHAGSGGKVKHOCERRFYRDKVRRHVVHTGKQVLCYOC		920 CTCTGGGTCTGCTCACCAATGATAAACTTTACCTTTGGGAAATGACATGCCCAATAGA	
AQRFGKRLHTRHMKSHNOELLKVKTEPVDLPDFTCMNSVIPKIDELLPVMSLPSE		961 TATGGACATGTTTCATCCCTCTCACACCTTTCTTCAAATATCCGTTTCAGTTTCTACCTC	
LLSKPFTNTLOLNLVNTFQSMQSSGSAHQMTTLPLGNTCFDMDVAPSHHLAFKC		980 TATGGATGCTGTTCACCCCTCTCATCATCTTCTGTTTCAAATGCCATTTCAGTTCTACCTC	
PFSSTYALSIPEKQPLKGETESYLMELQGGAPSSQDPSASSKLGLEPQSGPDD		1021 ATATGCAATTTCTATTTCCTGAAAAGAACAGACCATTAAGAGGGGAAATGAGATTACCT	
GAGDLSKSSSISIDPLSTPALDQSFNFIFPLNGPPYNTLSVSLGMSYQEBHNS		1040 ATACGCAATCTCTATTCCTGAAAAGAACAGACCATTAAGAGGGGAAATGAGATTACCT	
SVSQPTQDQIDQDPANTVGLSLHSLSAFTSSLSSTTLPRFHQAFQ"		1081 GATGAGTTTACAGGTGGCGTCCCTCTTCATCCCAAGATTTCTCAAGCATCTCATCATC	
ORIGIN		1100 GATGAGCTTACAGGTGGTGACCATCTTCATCCCAAGATTTCTCCAGCATCG---TCATC	
Query Match		1141 TAAGCTAGGTTGGATCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCT	
Best Local Similarity 84.3%; Score 1265.8; DB 10; Length 1575;		1157 TAAGCTAGGTTAGAGCCTCAGAGTGGTGGTCCAGATGATGGTGGGACCTCTCCCT	
Matches 1361; Conservative 0; Mismatches 137; Indels 3; Gaps 1;		1201 ATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACAGCATTTGGATTTCTCA	
1 GATGCCACTGTGATTCCTGTGTGATTTGTGAGAAGTAAAGAGATACCCAGAAAGTCCCTTC		1217 GTCAAGAGCTCTATCTCTATCAGTGACCCCTCAGCACAGCATTTGGATTTCTCTCA	
20 GATGCCACTGTGATTCCTGTGTGATTTGTGAGAAGTAAAGAGATACCCAGAAAGCCCTTC		1261 GTTCTTAATTTTATACCTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCT	
61 AGGGAAGAGTAAAGCTGTGAAACCAACCAAGAAACAACTTTCTTCCCTGCAACTGTGTA		1277 GTTCTTAATTTTATACCTTTAAATGGTCTCCCTATAATCTCTATCAGTGGGAGCT	
80 AGGGAAGAGTAAAGCTGTGAAACCAACCAAGAAACAACTTTCTTCCCTGCAACTGTGTA		1321 TGGATGAGCTATTTCCAGGAGAAAGACATTTCTTGTTCAGAGTCCCCACACAAAC	
121 CAAGGCCCTTTAAAGCTGTGAAACCAACCAAGAAACAACTTTCTTCCCTGCAACTGTGTA		1337 TGGATGAGCTATTTCCAGGAGAAAGACATTTCTTGTTCAGAGTCCCCACACAAAC	
140 CAAGGCCCTTTAAAGCTGTGAAACCAACCAAGAAACAACTTTCTTCCCTGCAACTGTGTA		1381 ACAGGATCTTCAGATCTCTGCAACACTATAGAGCTTGGGCTTCTGCTCTCAGTGGGAGCT	
181 GCCCTCAAGTGCATACCAACAGACTGCACCAAGGCTTTGTTTCTTAAGTACAAATTA		1397 ACAGGATCTTCAGATCTCTGCAACACTATAGAGCTTGGGCTTCTGCTCTCAGTGGGAGCT	
200 GCCCTCAAGTGCATACCAACAGACTGCACCAAGGCTTTGTTTCTTAAGTACAAATTA		1441 AGCTTCCAGCAGAGTTTAAAGCAAGTACACCTCCAGCTTCCAGTTCCTCAAGCTTTTCA	
241 AAGCCATGCTACTACTCTCTCTGAGAAACCAACCAAGTGTAAATTTCTTGAAGAAAT		1457 AGCTTCCAGCAGAGTTTAAAGCAAGTACACCTCCAGCTTCCAGTTCCTCAAGCTTTTCA	
260 AAGCCATGCTACTACTCTCTCTGAGAAACCAACCAAGTGTAAATTTCTTGAAGAAAT		1501 G 1501	
301 GTTTCACCGGAAGATCATCTGGAAGATCACTCCATACACACAGCCCTTAAAGCTCACTT		1517 G 1517	
320 GTTTCACCGGAAGATCATCTGGAAGATCACTCCATACACACAGCCCTTAAAGCTCACTT		RESULT 6	
361 GTTTCACCGGAAGATCATCTGGAAGATCACTCCATACACACAGCCCTTAAAGCTCACTT		AY574219	
380 GTTTCACCGGAAGATCATCTGGAAGATCACTCCATACACACAGCCCTTAAAGCTCACTT		LOCUS	
421 GGCCTTGCATGCCGCAACAGTGTGACCTCACTGTAAGGTATGTTGCAAACTTTTGA		DEFINITION	
440 GGCCTTGCATGCCGCAACAGTGTGACCTCACTGTAAGGTATGTTGCAAACTTTTGA		Mus musculus pleomorphic adenoma of the salivary gland 1 (Plagl1)	
481 AAGCAGGGAGTGTCTTCTGAGCAGCTTAAATCTCATGAGGCAAGTCTGCTGGTGGGT		mRNA, complete cds; alternatively spliced.	
500 AAGCAGGGAGTGTCTTCTGAGCAGCTTAAATCTCATGAGGCAAGTCTGCTGGTGGGT		AY574219.1 GI:46254398	
541 TAAAGAAAAGACACAGTGTGCAACATTTGTGATCGCCGCTTCTACACCGGAAGATGT		Mus musculus (house mouse)	
560 TAAAGAAAAGACACAGTGTGCAACATTTGTGATCGCCGCTTCTACACCGGAAGATGT		Mus musculus	
601 CCGGAGACATGCTGGTGTGACACCTGGAAGAAAGACTTCTCTGTCTGCTGCTGCTGCTG		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
620 CCGGAGACATGCTGGTGTGACACCTGGAAGAAAGACTTCTCTGTCTGCTGCTGCTGCTG		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
661 GAGATTGGGCGGAAGGATCACTGCTGCAATATGAAGAGAGTCAATCAAGAGCT		1 (bases 1 to 3161)	
680 GAGATTGGGCGGAAGGATCACTGCTGCAATATGAAGAGAGTCAATCAAGAGCT		Castilla, L.H.	
721 TCTGAAGGTCAAAACAGAACCTGATTTCTTGAACCCATTTACTGCAATGTCTGT		Direct Submission	
740 TCTGAAGGTCAAAACAGAACCTGATTTCTTGAACCCATTTACTGCAATGTCTGT		Submitted (16-MAR-2004) Program in Gene Function and Expression,	
781 GCCTATAAAGAGAGCTCTTCCCGTGTGATGCTTCTTCCAGTGAAGTCTTATCAAA		University of Massachusetts Medical School, 364 Plantation Street,	
800 GCCTATAAAGAGAGCTCTTCCCGTGTGATGCTTCTTCCAGTGAAGTCTTATCAAA		Worcester, MA 01605, USA	
841 GCCATTCAAAACACTTTTGCAGTTAACTCTTAAACACTCAATTCAGTCCATGACAGAG		Location/Qualifiers	
		1..3161	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="129/SvEv"	
		/db_xref="taxon:10090"	
		/chromosome="4"	


```
/tissue_type="whole embryo"
/dev_stage="14.5 dpc"
1. 3161
/gene="p1ag1"
581..2080
/gene="p1ag1"
/notes="transcription factor with 7 C2H2 zinc fingers;
alternatively spliced; contains exon 2b"
/codon_start=1
/product="pleomorphic adenoma of the salivary gland 1"
/db_xref="GI:46254393"
/translation="MATVIPGDLSEVRDQKAPSGKRRKRGSKPRKPPPCOLCDKAFN
SVEKLKVSFHSHTGERPKYKQDCTKAFVSKYKILORHMAHSPKTHKONYCKRMFH
RKDLKHLHLDHNKFTKBECKGKSYNTKLFGRHLALHAHRSGLDTCVKCLQNEP
STGVLEHLKHAGSKSGVKEKHQCEKRRFYTRKDRVRHVVHVPKGRKDFIQCYC
AORERKDLHTRHMKSHNQELLVKTEPVDLPDFTCNMSPVIKDELLPVSHPFSE
LLSKPFTNLOLNLYNTPQSMQSSGSAHOMITLPLGMCPIIDMDVHPHHLAFKC
PFSSTVAISIPKEQPLKGRLESLVLMELQSGAPSSODSPASSKLGLPQSGSPDD
GAGDLSUSKSSISDPLSTPALDPSQLPFPNPLPPLNGPPYNTLSVGLMSYSQBEAHS
SVSQLPTQDLPDPANTVGLSSLSHLSAAFTSSLSSTLPRFHQAQ"
```

ORIGIN

Query Match	84.3%; Score 1265.8; DB 10; Length 3161;
Best Local Similarity	90.7%; Pred. No. 0;
Matches 1361; Conservative	0; Mismatches 137; Indels 3; Gaps 1;

1 GATGGCCACTGTCATTCCTCGTGTGATTGTGAGAGTAAGAGATACCAGAAAGTCCCTTC 60

580 GATGGCCACTGTCATTCCTCGTGTGATTGTGAGAGTAAGAGATACCAGAAAGTCCCTTC 639

61 AGGGAACGTGACGCTGTGTAACCAACCAAGAAAGAACTTCCTTGCCAACTGTGTGA 120

640 AGGGAACGTGACGCTGTGTAACCAACCAAGAAAGAACTTCCTTGCCAACTGTGTGA 639

121 CAAGGCTTTAAAGTGTGAGAGAAATTAAGGTTCACTCTCTCTCACACAGGAGAG 180

700 CAAGGCTTTAAAGTGTGAGAGAAATTAAGGTTCACTCTCTCTCACACAGGAGAG 759

181 GCCTCAAGTGCATACAAACAAGACTGCACAGGCGCTTTGTTCTAAGTACAAATTACA 240

760 GCCTCAAGTGCATACAAACAAGACTGCACAGGCGCTTTGTTCTAAGTACAAATTACA 819

241 AAGGCATCGCTACTCATCTCTGAGAAACCAAGTGAATTAATGTGAGAAAT 300

820 AAGGCATCGCTACTCATCTCTGAGAAACCAAGTGAATTAATGTGAGAAAT 879

301 GTTTCACGGGAAGATCATCTGAAGATCACTCTCATACACACGACCTTAACAAAGAG 360

880 GTTTCACGGGAAGACCATCTGAAGATCACTCTCATACACACGACCTTAACAAAGAG 939

361 GTTTCACGGGAAGATCTGCAAGAACTACAATACCAAGCTTGATTTAAAGTCACTT 420

940 CTTTAAATGCAAGAGTGTGCAAGAGCTACAATACCAAGCTTGATTTAAAGTCACTT 999

421 GGCCTTGATCGGCAACAAAGTGTGACCTCACCTGAAGTATGTTTGCAAACTTTGA 480

1000 GGCCTTGATCGGCAACAAAGTGTGACCTCACCTGAAGTATGTTTGCAAACTTTGA 1059

481 AAGCACGGAGTGTCTTCTGGAGCCTTAAATCTCATGACGCAAGTCTGTGTTGGGT 540

1060 GAGCACAGTGTGTCTTCTAGACACCTGAAGATCACTCTCATACACGCGGCAAGTCA 1119

541 TAAAGAAAAAGACCAAGTGTGCAACATTTGATCGCGGTCTTCAACCGGAAGAGTCT 600

1120 GAAGGAGAAAAAGCACCATGTGACACTGCGAAGCGAGTTCTACACCGGAAGAGTCT 1179

601 CCGGAGACACATGTTGTGTCACACTGGAAGAGAGTCTTCTCTGTGATTTGTGCAACA 660

1180 CCGGAGACACATGTTGTGTCACACAGGAGAGAGAGTCTTCTCTGTGATTTGTGCAACA 1239

661 GAGATTGGCGGGAAGAGTCACTGCTGATGATGAAGAGAGTCAAAATCAAGAGCT 720

1240 GAGATTGGAGAAAGGATCACCTCACTCGACACATGAAGAGAGTCAAAATCAAGAGCT 1239

721 TCTCAAGGTCAAAACAGAACAGTGGATTTTCCTTGACCCATTTACCTGCAATGTGTGT 780

1300 TCTGAAGGTCAAAACAGAACAGTGGATTTTCCTTGACCCATTTACCTGCAACATGTGT 1359

781 GCCTATAAAGACAGAGCTCTTCGGTGTGATGTCCTTACCTTCCAGTGAATTTATCAAA 840

1360 GCCTATAAAGAGTAACTCTCGCGGTGATGTCCTTACCTTCCAGTGAATTTATCAAA 1419

841 GCCATTACAAACACTTTTGACAGTTAAACCTCTCAACAACTCAATTTAGTCAATCAGAG 900

1420 GCCATTACAAACACTTTTGACAGTTAAACCTCTCAACAACTCAATTTAGTCAATCAGAG 1479

901 CTCGGAGTCTCCCAACAAATGATCAAACTTTTACCTTTGGAAATGATGATGCCCATAGA 960

1480 CTCGGGTCTGCTCACCAATGATAAACAACTTTTACCTTTGGAAATGATGATGCCCATAGA 1539

961 TATGACACTCTGTCATCTCCCTCTCACCACTTTCTTTCAAATATCCGTTCAAGTTCACCTC 1020

1540 TATGATGCTGTTTACCCTCTCTCATCTCTTCTTCAATGCCCATTTCAGTTTACCTC 1599

1021 ATATCAATTTCTATTCTCTGAAAGAAAGACAGCCATTAAAGGGGAAATTTAGAGTTACCT 1080

1600 ATACCAATCTCTATTCTCTGAAAGAAAGACAGCCATTAAAGGGGAAATTTAGAGTTACCT 1659

1081 GATGAGATTCAAGTGGGTGCTCTCTCTTCAATCCCAAGATTCTCAAGCATCTGTCATCATC 1140

1660 GATGAGATTCAAGTGGGTGCTCTCTCTTCAATCCCAAGATTCTCTCAGCATCTG---TCATC 1716

1141 TAAGCTAGGTTTGGATCTCTCAGATTGGGTCTCTAGATGATGTCAGGAGACCTCTCCCT 1200

1717 TAAGCTAGGTTTGGATCTCTCAGATTGGGTCTCTAGATGATGTCAGGAGACCTCTCCCT 1776

1201 ATCAAAAGCTCTATCTCCATCAGTCAACCCCTTAACACACACAGCATTTGATTTTCTCA 1260

1777 GTCAAAGAGCTCTATCTCTATCAGTCAACCCCTCAGCACACAGCATTTGATTTCTCTCA 1836

1261 GTTGTAAATTTATACCTTTAAATGCTCTCCCTATAATCTCTATACAGTGGGAGCT 1320

1837 GTTGTAAATTTATACCTTTAAATGCTCTCCCTATAATCTCTATACAGTGGGAGCT 1896

1321 TGAATGAGCTTTTCCAGGAAAGACACATTTCTTGTGTTTCCAGCTCCCAACAAAC 1380

1897 TGGGATGAGTATTTCCAGGAAAGACACATTTCTTGTCTCTCAGCTGCCACACAAAC 1956

1381 ACAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGACTCACTGTGAGC 1440

1957 ACAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGACTCACTGTGAGC 2016

1441 AGCTTTACAGGATTTTAAAGCAAGTACCACTCCCTCCAGCTTTTCCATCAAGCTTTTCA 1500

2017 AGCTTTACAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGACTCACTGTGAGC 2076

1501 G 1501

2077 G 2077

RESULT 7

AC107952/c 142102 bp DNA linear PRI 23-JUL-2002

LOCUS Homo sapiens chromosome 8, clone RP11-140I16, complete sequence.

DEFINITION AC107952

ACCESSION AC107952

VERSION AC107952.5 GI:21930261

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 142102)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 8, clone RP11-140I16

[illegible]

540	TTAAAGAAAAAAGCAACAGTGGGAACATTTGTGATCGCGGTTCTACACCCGAAGGATG	599
144485	TGAAAGAGAAAAAGCACCAGTGTGAGCACTCGAGCGCAGGTTCTACACCCGGAGGATG	144544
600	TCCGAGAGCACATGTTGGTGGCACCTGGAAGAAAGGACTTCCTCTGTGAGTATTGTGGAC	659
144545	TTCCGAGACACATGTTGGTGGCACCGGAGAAAGGACTTCCTCTGTGAGTATTGTGGAC	144604
660	AGAGATTTCGGGCGAAAGGATCACCTGATCTCGACATATGAGAGAGTACATCAAGAGC	719
144605	AGGATTTTGGACGAAAGGATCACCTCACATGAGAGAGTACATCAAGAGC	144664
720	TTCTGAAGGTCAAAACAGAACAGTGGATTTCCTTGACCCATTACCTGCAATGTGTCTG	779
144665	TTCTGAAGGTCAAAACAGAACAGTGGATTTCCTTGACCCATTACCTGCAATGTGTCTG	144724
780	TGCGCTATAAAGACGAGCTCCTTCGCGGTGATGTCCTTACCTGCCAGTGAATCTGTTATCAA	839
144725	TGCGCTATAAAGATGAGTCTCTCGCGGTGATGTCCTTACCTGCCAGTGAATCTGTTATCAA	144784
840	AGCCATTTCACAAACACTTTGAGTTAAACCTCTACACACTCCATTTTCAGTCCATCGACA	899
144785	AGCCATTTCACAAACACTTTGAGTTAAACCTCTACACACTCCATTTTCAGTCCATCGACA	144844
900	GCTCGGATCTGCCACCAATGATCAAACTTACCTTTGGGAATGACATGCCCAATAG	959
144845	GCTCTGGGTCTACTCACAAATGATCAAACTTACCTTTGGGAATGACATGCCCAATAG	144904
960	ATATGGACACTGTTTCATCCCTCTCACACCTTTCTTCAAATATCCGTTTCAGTTCTACCT	1019
144905	ATATGGACACTGTTTCATCCCTCTCACACCTTTCTTCAAATATCCGTTTCAGTTCTACCT	144964
1020	CATATGCAATTTCTATTCTGAAAGAGACAGCCATTAAAGGGGGAAATTTGAGAGTTACC	1079
144965	CATACGCAATCTCTATTCCCGAGAAAGAACAGCCCTTAAGAGGGGAAATCGAGAGTTATC	145024
1080	TCATGGAGTTACAAGGTGGCGGCCCTCTTCATCCCAAGATTCTCAAGCATGTCATCAT	1139
145025	TCATGGAGTTACAAGGTGGCGGCCCTCTTCATCCCAAGATTCTCAAGCATGTCATCAT	145081
1140	CTAAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGAGACCTCTCCC	1199
145082	CTAAGTTAGGGTTAGAGCCTCAGAGTGGCTCCCGCAGACGATGGTCTGGGAGCTCTCCC	145141
1200	TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACACGACATTTGGATTTTCTC	1259
145142	TGTCAAAGAGCTCTATCTCTATCAGTGACCCCTGAAACACACGACATTTGGATTTTCTC	145201
1260	AGTTGTTTAAATTTTCAATACCTTTAAATGGTCTCCCTCTATAATCCTCTACGTGGGAGCC	1319
145202	AGTTGTTTCAATTTTCAATACCTTTAAATGGTCTCCCTCTATAATCCTCTACGTGGGAGCC	145261
1320	TTGGAATGAGCTAATTCAGGAAAGAGCACATTTCTGTTTCCGAGCTCCCCACACAAA	1379
145262	TTGGATGAGCTAATTCAGGAAAGAGCACATTTCTGTTTCCGAGCTCCCCACACAAA	145321
1380	CACAGGATCTTCAGGATCTTCGAAACACTATAGGGTTGGGTCTCTGCACTCACTGTGAG	1439
145322	CACAGGATCTTCAGGATCTTCGAAACACTATAGGGTTGGGTCTCTGCACTCACTGTGAG	145381
1440	CAGCTTTTCACAGCAGTTTAAAGCAAGATACACCTTCCACGTTTCCATCAGCTTTTC	1499
145382	CAGCTTTTCACAGCAGTTTAAAGCAAGATACACCTTCCACGTTTCCATCAGCTTTTC	145441
1500	AG 1501	
145442	AG 145443	

RESULT 9

AC097274/c

LOCUS

DEFINITION

182102 bp

182102 bp

182102 bp

182102 bp

linear

linear

linear

linear

HTG 18-JAN-2002

AC097274

Mus musculus clone RP23-59B17, WORKING DRAFT SEQUENCE. 8 unordered

RESULT	9
AC097274/c	
LOCUS	182102 bp DNA linear HTG 18-JAN-2002
DEFINITION	Mus musculus clone RP23-59B17, WORKING DRAFT SEQUENCE, 8 unordered

```

pieces.
AC097274      GI:18201773
AC097274.5    GI:18201773
VERSION       HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS      Mus musculus (house mouse)
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
-----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
On Jan 18, 2002 this sequence version replaced gi:16973738.
-----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
-----
Center project name: RP23-59B17
Center clone name: RP23-59B17
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 48370: contig of 48370 bp in length
* 48371 43082: gap of unknown length
* 49083 95033: contig of 45951 bp in length
* 95034 95745: gap of unknown length
* 95746 129069: contig of 33324 bp in length
* 129070 129781: gap of unknown length
* 129782 150082: contig of 20301 bp in length
* 150083 150793: gap of unknown length
* 150794 162439: contig of 11646 bp in length
* 162440 163150: gap of unknown length
* 163151 172099: contig of 8949 bp in length
* 172100 172810: gap of unknown length
* 172811 178596: contig of 5786 bp in length
* 178597 179307: gap of unknown length
* 179308 182102: contig of 2795 bp in length.
*
* Location/Qualifiers
* 1. .182102
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP23-59B17"
*
FEATURES
source

```

[illegible]

Db 60713 TGTTTACCGGAAAGACCATCTGAAGATACCTCCATACATGACATGACCCCAACAAAGAGA 60654

QY 360 CGTTTAAAGTCGGAAGATGCGAAGAACTACAATACCAAGCTTGGATTTAAAGCTCACT 419

Db 60653 CCTTTAAATCGAAGAGTGTGGAAGAGCTACAATACCAAGCTTGGTTTAAAGCGACACT 60594

QY 420 TGGCCTTGCATGCGGCAACAAAGTGTGACCTGACCTGTAAGGTATGTTTGCACAACTTTTG 479

Db 60593 TGGCCTTGCATGCTGCCACAGTGGTGACCTGCACTGCAAGGTGTGTTTGCAGAAATTTTG 60534

QY 480 AAAGCAGGAGTGTCTTCTGGAGCACCTTAAATCTCATGACGAGCAAGTGTCTGTGGGG 539

Db 60533 AGAGCACAGTGTGCTCTCTAGACCACTGAAATCTCAGCGGCAAGTCACTGTGGGGTG 60474

QY 540 TTAAAGAAAAGACACAGTGGCGAATGTTGATCGCCGGTTCTTACACCCGAAAGGATG 599

Db 60473 TGAAGAGAAAAGACCAATGTGAGCACTGCGAAGCGAGGTCTTACACCCGGAAGATG 60414

QY 600 TCCGAGACACATGTGTGGTGCACTGGAAGAAAGCACTTCTCTGTCTGATTTGTCAC 659

Db 60413 TCCGAGACACATGTGTGTAACACAGGAAGAAAGCACTTCTCTGTCTGATTTGTCAC 60354

QY 660 AGAGATTTGGCGAAGAGTCACTGACTCGACATATGAAGAAAGTCAATCAATCAAGAC 719

Db 60353 AGAGATTTGGCGAAGAGTCACTGACTCGACATATGAAGAAAGTCAATCAATCAAGAC 60294

QY 720 TTCTGAAGTCAAAACAGAACAGTGGATTTCTTGAACCCATTTACCTGCAATGTGTCTG 779

Db 60293 TTCTGAAGTCAAAACAGAACAGTGGATTTCTTGAACCCATTTACCTGCAATGTGTCTG 60234

QY 780 TGCCTATAAAGACAGCTCTTCCGGTGATGTCCTTACCTTCAGTGAATGTTTATCAA 839

Db 60233 TGCCTATAAAGATGAATCTCTCGGGTCATGTCCTTACCTTCAGTGAATGTTTATCAA 60174

QY 840 AGCATTCACAAACATTTGAGTTAAACCTCTCAACACATCCATTTCAAGTCCATGAGA 899

Db 60173 AGCATTCACAAACATTTGAGTTAAACCTCTCAACACATCCATTTCAAGTCCATGAGA 60114

QY 900 GCTCGGATCTGCCCAACAAATGATCAACATTTACCTTTGGGAATGACATGCCCAATAG 959

Db 60113 GCTCGGATCTGCCCAACAAATGATCAACATTTACCTTTGGGAATGACATGCCCAATAG 60054

QY 960 ATATGGACATGTTTATCCCTCTCACCACCTTCTTCAATATCCGTTCAAGTTCTACCT 1019

Db 60053 ATATGGATGTTTATCCCTCTCACCACCTTCTTCAATATCCGTTCAAGTTCTACCT 59994

QY 1020 CATATGCAATTTCTATTCCTGAAAGAAACAGCCATTAAGGGGGAATTCAGAGTTACC 1079

Db 59993 CATATGCAATTTCTATTCCTGAAAGAAACAGCCATTAAGGGGGAATTCAGAGTTACC 59934

QY 1080 TGATGGATTCAGAGTGGGTCCTCTTCAATCCCAAGATTCCTCAAGCATCGTCATCAT 1139

Db 59933 TGATGGATTCAGAGTGGGTCCTCTTCAATCCCAAGATTCCTCAAGCATCGTCATCAT 59877

QY 1140 CTAAGTAGGTTGGATCTCAGATGGGTCCTAGATGATGTCGAGAGACCTCTCC 1199

Db 59876 CTAAGTAGGTTGGATCTCAGATGGGTCCTAGATGATGTCGAGAGACCTCTCC 59817

QY 1200 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTAAACACACAGCATTTGGATTTTCTC 1259

Db 59816 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTAAACACACAGCATTTGGATTTTCTC 59757

QY 1260 AGTTGTTTAAATTCATACCTTTAAATGGTCCCTCTATATCTCTATCAGTGGGAGCC 1319

Db 59756 AGTTGTTTAAATTCATACCTTTAAATGGTCCCTCTATATCTCTATCAGTGGGAGCC 59697

QY 1320 TTGGATGAGCTTATCCAGGAGAGACATCTTCTGTTTCCAGCTCCCAACACAAA 1379

Db 59696 TTGGATGAGCTTATCCAGGAGAGACATCTTCTGTTTCCAGCTCCCAACACAAA 59637

QY 1380 CACAGATCTTCAGGATCTGCAAAACACTATAGGCTTGGTCTCTGCACTCACTGTGAC 1439

Db 59636 CACAGATCTTCAGGATCTGCAAAACACTATAGGCTTGGTCTCTGCACTCACTGTGAC 59577

QY 1440 CAGCTTTACCAAGAGCTTTAAAGCAAGTACCACTCCACGTTTCCATCAAGCTTTTC 1499

Db 59576 CAGCTTTACCAAGAGCTTTAAAGCAAGTACCACTCCACGTTTCCATCAAGCTTTTC 59517

QY 1500 AG 1501

Db 59516 AG 59515

RESULT 10

AL807387/c

DEFINITION

Mouse DNA sequence from clone RP23-24J10 on chromosome 4, complete

sequence.

AL807387

ACCESSION

AL807387

VERSION

AL807387.10

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Harrison.E.

REFERENCE

1 (bases 1 to 241148)

AUTHORS

Direct Submission

JOURNAL

Submitted (03-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On May 3, 2003 this sequence version replaced gi:26788130.

Sequence from the Mouse Genome Sequencing Consortium whole genome

shotgun may have been used to confirm this sequence. Sequence data

from the whole genome shotgun alone has only been used where it has

a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep

from the RPI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

Location/Qualifiers

1..241148

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="4"

/clone_lib="RPI-23"

/clone="RP23-24J10"

/RPI-23

ORIGIN

Query Match

68.7%; Score 1031.6; DB 10; Length 241148;

Time Nov 9 09:06:21 2004

Best Local Similarity	89.1%;	pred. No. 6.3e-292;			
Matches 1125;	Conservative	0;	Mismatches 134;	Indels	3; Gaps 1;
QY	240	AAAGGCACATGGCTACATCTCTCTGAGAAACCCACAGTGTAAATATTGTGAGAAA	299		
Db	75206	AAAGGCACATGGCTACTCACTCTCTGAGAAACCCACAGTGTAAATATTGTGAGAAA	75147		
QY	300	TGTTTACCAGAAAGATCATCTGAGAAATCACTTCCATACACAGACCTTAACAAAGAGA	359		
Db	75146	TGTTTCCACGGAAGACCATCTGAGAAATCACTTCCATACACAGACCTTAACAAAGAGA	75087		
QY	360	CGTTTAAAGTGCAGAAATGTGCGAAGAACTACAATACCAAGCTTGGATTTAAACGTCAT	419		
Db	75086	CGTTTAAAGTGCAGAGGTGCGAAGAGCTACAATACCAAGCTTGGATTTAAAGCAGCAT	75027		
QY	420	TGGCCTTCATGCGCCACAAAGTGTGTGACCTCACTGTAAAGTATGTTTGCAAACTTTTG	479		
Db	75026	TGGCCTTCATGCTGCCACCAGTGTGTGACCTCACCTGTGAGTGTGTTTGCAGAAATTTG	74967		
QY	480	AAAGCACGGAGTGCTTCTGGAGCACCTTTAAATCTCATGACGGAAGTCTGTGTGGGG	539		
Db	74966	AGAGCACAGGTGTCTCTTAGAGCACCTGAAATCTCACGGCGCAAGTCATCTGGGGGTG	74907		
QY	540	TTTAAAGAAAAGACACAGTGTGGAACATGTGATCGCGGTCTTACACCCGAAAGCATG	599		
Db	74906	TGAAGGAGAAAAGACCAATGTGACACTGGAGCGAGGTCTTACACCGGAAGCATG	74847		
QY	600	TCCGGAGACATGGTGTGTGACACTGGAAGAAAGGACTTCCTCTGTCAGTATTGTGCAC	659		
Db	74846	TCCGGAGACATGGTGTGTGACACACAGGAGAAAGGACTTCCTCTGTCAGTACTGGCAC	74787		
QY	660	AGAGATTGGCGGAAAGGATCACTCACTGCAATGCAATATGAAGAAGTCACAANTCAAGAC	719		
Db	74786	AGAGATTGGACGAAAGGATCACTCACTGCAACATGAAGAAGTCACAANTCAGGAC	74727		
QY	720	TTCTGAAGTCAAAACAGACCAAGTGGATTTCCTTGACCCATTTACTGCAATGTGTCTG	779		
Db	74726	TTCTGAAGTCAAAACGGAACCAAGTAGATTTCCTGGACCATTTACTGCAACATGTCTG	74667		
QY	780	TGCTATATAAGACGAGCTCCTTCGCGTGTATGCTTTACCTTCCAGTGAATGTTATCAA	839		
Db	74666	TGCTATATAAGATGAATCCTTCGCGGTATGCTTTACCTTCCAGTGAATCTTGTCAA	74607		
QY	840	AGCAATTCACAAACATTTGGAGTTAACTCTACACACTCCATTTTCAGTCCATGAGA	899		
Db	74606	AGCAATTCACAAACATTTGCAGTTAACTCTATAACACTCCATTTTCAGTCCATGAGA	74547		
QY	900	GCTCGGATTCGCCACCAATATGATCAAACTTTTACCTTTGGGAATGATGCCCCATAG	959		
Db	74546	GCTCTGGGTCTGCTACCAATATGATCAAACTTTTACCTTTGGGAATGATGCCCCATAG	74487		
QY	960	ATATGACATGTTTCATCCCTCTCACACCTTTCTTCAAATATCCGTTTCAGTTCTACT	1019		
Db	74486	ATATGATGCTGTTTCAACCCCTCTCATCATCTTGTCTCAAATGCCCCATTCAGTTCTACT	74427		
QY	1020	CATATGCATTTCTTATTCCTGAAAGAAACAGCCATTAAGGGGGAAATTCGAGGTTACC	1079		
Db	74426	CATAGCCAAATCTCTAATCTGAAAGAAACAGCCATTAAGGGGGAAATTCGAGGTTACC	74367		
QY	1080	TGATGGAGTTACAAGTGGCGTCCCTCTTTCATCCCAAGATTCTCAAGATCGTCACTAT	1139		
Db	74366	TGATGGAGCTACAAGTGGTGGCAACATCTTCAATCCAGGATTCTCCAGCATCG---TCAT	74310		
QY	1140	CTAAGCTAGGTTTGAATCTCAGATTGGTCCCTAGATGATGGTGCAGAGACCTCTCCC	1199		
Db	74309	CTAAGCTAGGTTTAGAGCTCAGATGGGTCCGAGATGATGGTCTGGGGACCTCTCCC	74255		
QY	1200	TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAACACACAGCATTTGGATTTTCTC	1259		
Db	74249	TGTCAAAGAGCTCTATCTATCAGTGACCCCTCAGCACACCCAGCATTTGGATTTTCTC	74199		
QY	1260	AGTTGTTTAAATTCATACCTTTAAATGGTCCCTCTAATCTCTATCACTGATGGGGAGCC	1319		

Db	74189	AGTTGTTCAATTTTCATACCACTGAATGGTCCCTCTATTAACCCACTTTTCAGAGGAGCC	74191
Qy	1320	TTGGATGAGCTATTCCCGAGGAGAGCAGCATTTCTTGTGTTTCCAGCTCCCCACACAAA	13179
Db	74129	TTGGATGAGCTATTCCCGAGGAGAGCAGCATTTCTTGTCTCTCAGCTGCCACACAAA	74070
Qy	1380	CACAGGATCTTCAGGATCTCTGCAAAACACTATAGGGCTTGGGTTCTTGCACTCACTGTGAG	1439
Db	74069	CACAGGATCTTCAGGATCTCTGCAAAACACTATAGGGCTTGGGTTCTTGCACTCACTGTGAG	74010
Qy	1440	CAGCTTTCCACGAGCGTTTAAGCAGCAAGTACCACTCCCTCCAGTCTTCCATCAAGCTTTTC	1449
Db	74009	CAGCGTTCCACGAGCGCTGAGTCAAGCACTACCTTGCCCGCTTCCACGAGGCGTTTC	73950
Qy	1500	AG 1501	
Db	73949	AG 73948	

RESULT 11

AC123210/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC123210
Rattus norvegicus clone CH230-208L11, *** SEQUENCING IN PROGRESS
***, 5 unordered pieces.
AC123210.3 GI:23101071
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 244955)
Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anquiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W.,
Gunnaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,J.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenschwah,L., Louissegh,H., Lozardo,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackelmeah,O., Okuwonu,G., Olarunpasegun,A., Pal,S.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sned,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villabana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

REFERENCE AUTHORS TITLE JOURNAL

COMMENT

2 (bases 1 to 244955)

Worley, K. C.

Direct Submission

Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 244955)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 18, 2002 this sequence version replaced gi:21671673.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GRKE

Center clone name: CH230-208L11

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 214612 bases at least Q40

Consensus quality: 217681 bases at least Q30

Consensus quality: 219976 bases at least Q20

Estimated insert size: 240534; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 135745: contig of 135745 bp in length
* 135746 135845: gap of unknown length
* 135846 239680: contig of 103835 bp in length
* 239681 239780: gap of unknown length
* 239781 240825: contig of 1045 bp in length
* 240826 240925: gap of unknown length
* 240926 242465: contig of 1540 bp in length
* 242466 242565: gap of unknown length
* 242566 244955: contig of 2390 bp in length.

Location/Qualifiers

1. .244955

FEATURES source

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-208L11"

ORIGIN
Query Match      64.2%; Score 964.2; DB 2; Length 244955;
Best Local Similarity 88.9%; Pred. No. 4.3e-272;
Matches 1054; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 317 CATCTGAAGAATCACCTCCATACACAGCACCCTACAAAGAGACGTTTAAAGTCGGAAGAA 376
      |||
Db 223660 CATCTGAAGAATCACCTCCATACACAGCACCCTACAAAGAGACGTTTAAAGTCGGAAGAG 223601

QY 377 TGTGGCAAGAACTACAATACCAAGCTTGGAATTAACAGCTCACTTGGCCCTTGATGCCGCA 436
      |||
Db 223600 TGTGGCAAGAGTACAACACCAAGCTTGGTTTAAGCGACACTTGGCCCTTGACCGTGGC 223541

QY 437 ACAAGTGGTGACCTCACCTGTAAAGTATGTTTGGAAAATTTTGAAGCAACGGAGTGT 496
      |||
Db 223540 ACCAGTGGCGACCTCACCTGTAAAGTATGTTTGGCAGACTTTTGAGAGCACAGGTGTGCTC 223481

QY 497 CTGAGACACCTTAAATCTCATGSCAGGCAAGTCCTCTGCGGCGGTAAAGAGAAAAAGCAC 556
      |||
Db 223480 CTAGAGACCTTGAATCTCAGCAGGCAAGTCCTCTGCGGCGGTAAAGAGAAAAAGCAC 223421

QY 557 CAGTGCAGAACATGATGATGCGCGGTCTTACACCCGAAAGAGATGTCGCGAGACACATGTTG 616
      |||
Db 223420 CAGTGTGAGCACTCGAGCGCAGGTTCTTACACCCGAAAGAGATGTCGCGAGACACATGTTG 223361

QY 617 GTGCACACTGGAAGAAAGGACTTCTCTGTGATGTTGTGCACAGAGATTTGGGCGAAG 676
      |||
Db 223360 GTGCACAGGGAAGAAAGGACTTCTCTGTGATGTTGTGCACAGAGATTTGGGCGAAG 223301

QY 677 GATCACCTGATCGACATATGAAGAAGTCAACATCAAGAGCTTCTGAAGGTCAAAACA 736
      |||
Db 223300 GATCACCTGATCGACATATGAAGAAGTCAACATCAAGAGCTTCTGAAGGTCAAAACA 223241

QY 737 GAACAGTGAATTCCTTGACCCCAATTTACCTGCAATGTGTCTGTCCTATATAAGACGAG 796
      |||
Db 223240 GAACAGTGAATTCCTTGACCCCAATTTACCTGCAATGTGTCTGTCCTATATAAGAGTAG 223181

QY 797 CTCCTCCGGTGATGTCCTTACCTTCAGTGAATGTTATCAAGCCCAATTCACAAACACT 856
      |||
Db 223180 CTCCTCCGGTGATGTCCTTACCTTCAGTGAATGTTATCAAGCCCAATTCACAAACACT 223121

QY 857 TTGCAGTTAAACCTCTACAACACTTCCATTTTCAGTCCATGAGAGCTCGGATCTGCCAC 916
      |||
Db 223120 TTGCAGTTAAACCTCTACAACACTTCCATTTTCAGTCCATGAGAGCTCGGATCTGCCAC 223061

QY 917 CAAATGATCAAACTTTACCTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCAT 976
      |||
Db 223060 CAAATGATCAAACTTTACCTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCAT 223001

QY 977 CCTCTCACCACTTTCTTCAAAATCCGTTACGTTTACCTCATATGCAATTTCTATT 1036
      |||
Db 223000 CCTCTCATCATCTTGCCTTTCAAAATCCGTTTACGTTTACCTCATATGCAATTTCTATT 222941

QY 1037 CCTGAAAAAGAACAGACCATTTAAAGGGGAAATGAGAGTTACTGATGAGGTTACAGGT 1096
      |||
Db 222940 CCGGAGAAAGAACAGACCATTTAAAGGGGAAATGAGAGTTACTGATGAGGTTACAGGT 222881

QY 1097 GCGGTGCTCTTCATCCCAAGATTTCTCAAGCATGCTCATCATCTAAGCTAGGTTGGAT 1156
      |||
Db 222880 GCGGACCATCTTCATCCCAAGATTTCTCAAGCATGCTCATCATCTAAGCTAGGTTGGAT 222824

QY 1157 CCTCAGATGGGTCCCTAGATGATGGTGAGAGACCTCTCCCTATCCAAAAGCTCTATC 1216
      |||
Db 222823 CCTCAGATGGGTCCCTAGATGATGGTGAGAGACCTCTCCCTATCCAAAAGCTCTATC 222764

QY 1217 TCATCAGTGACCCCTTAAACACACAGCATGGATTTTCTCAGTTGTTTAAATTCATA 1276
      |||
Db 222763 TCTATCAGTGACCCCTTAAACACACAGCATGGATTTTCTCAGTTGTTTAAATTCATA 222704

```


Blakesley, R.W., Bouffard, G.G., Breen, K., Guan, X., Gupta, J., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Karfins, P., Kwong, P., Laric, P., Legaspi, R., Hansen, N., Ho, S.-L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., Maduro, Q.L., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., McDowell, J., Voigt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Tsurgon, C., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov> Series: IRAL Plate: 39 Row: f Column: 4 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6031195.

location/Qualifiers
1. .1820
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29597 IMAGE:4764127"
/tissue_type="Primary B-Cells from Tonsils"
/clone_lib="NIH MGC_48"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. .1820
/gene="PLAGL2"
/db_xref="LocusID:5326"
/db_xref="MIM:604866"
204. .1694
/gene="PLAGL2"
/codon_start=1
/product="pleiomorphic adenoma gene-like 2"
/protein_id="AAH23655.1"
/db_xref="GI:23271120"
/db_xref="LocusID:5326"
/db_xref="MIM:604866"
/translation="MTFTFTSVPPWTDQAKVEEVGKLVPRPRREARSQVCKCEI
STPSPNGKLRPHSLPQEPQPCQHLCHGKAFASKLYRHMATHAASGGLSKV
CDMEHRKDLRNHLQTHDPNKEALHSCGKYNKTLGYRRLHMAHSAAGSLCKV
CIQTFSTQALLHLKASRRVAGGAKKKHPCDCHDRFRYTRKDVRRHLVHTGRK
FLCQYCAQRFGRKDLHTRVKKSHOELLKIKTEPDMGLLSLSSAEVPLESPSS
CNASDVMTGKAPFGMLPMGYGAHPTMPSITGVPSHLVHTLPMGMYPLESPSS
PQALPKYQIGTSYLPDLKPEVDSFLAELPGLSLSSAEVPLESPSS
EALLAKSPANLSEALCAANVDFSHLGLFLNLPNCPNPGATGGLVNGYSQAEQPL
TTLQAPQSPGAGGLNFGPLHSLFPVFTSGLSSTLPRFHQAFQ"

ORIGIN
Query Match 26.0%; Score 389.8; DB 9; Length 1820;
Best Local Similarity 58.2%; Pred. No. 5e-103;
Matches 823; Conservative 0; Mismatches 542; Indels 48; Gaps 6;
Qy 98 AACTTTCTTCCCACTGTGTGACAGAGCCCTTTAACAGTGTGAGAAATTAAGGTTCCAC 157
Db 318 AAGTGCCAAATGTGAATTTTCGGGAACACCTTTCTCAATGGGAGAGAGCTGAGCCCTC 377
Qy 158 TCCTACTCTCACAGAGAGAGAGCCCTTCAAGTGCATACAAAGACTGCACCAAGGCC 217
Db 378 AGCTCCCGACACAGAGAGAGACCATATAGCTGCCCTCAGCTGCACTGTGGCAAGGCT 437
Qy 218 TTTGTTTCTTAAGTACAAATTAACAAGGCACATGGCTACTTCTCTCTGAGAAACCCAC 277
Db 438 TTTGTTTCTTAAGTACAAATTAACAAGGCACATGGCTACTTCTCTGAGTGGGAGAGAGCC 497
Qy 278 AAGTGTAAATTTGTGAGAAATTTTACCGGAAAGATCATCTGAGAAATCACTCCATCCAT 337
Db 498 CAGTGTATGTACTGTGTAAGATGTTTCCACGAGGAGACCATCTGCGGAAACCATCTGAG 557
Qy 338 ACACAGAGCCCTTACAAAGAGAGAGCTTTTAAGTGGGAGAAATGTGGCAAGACTTACAAAT 397
Db 558 ACCATGATCTTACAAAGAGAGCCCTTCCACTGCTGAGTGGGTAAGAAATTAACAATAG 617
Qy 398 AAGCTTGGATTAAACGTCACCTTGGCCCTTGATCGCGCAACAGAGTGTGACCTACCTCT 457
Db 618 AAGCTGGGCTACCGCGGCCACCTGGGCCATGATGTCGCGAGCAGCGGTGACCTCAGCTGC 677

1277 CTTTAAATGCTCTCCTATATACCTCTATCAGTGGGAGCCCTTGAATGAGTATTC 1336
222703 CCATTAAATGCTCTCCTTACAAATCCGCTCTCCGTTGGGAAGCCCTTGGATGAGCTATTC 222644
Qy 1337 CAGGAGAGACATCTCTCTGTTTCCAGTCTCCACACAAACACAGATCTTCAGGAT 1396
Db 222643 CAGGAGAGGCGCATCTCTCTGTTTCTCAGTGTGCCACACAAACACAGATCTTCAGGAT 222584
Qy 1397 CTTGCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGAGCTTTTACAGGAGT 1456
Db 222583 CTTGCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGAGAGCTTTTACAGGAGT 222524
Qy 1457 TTAAGCAAGATGACACCTCCAGCTTCCATCAAGCTTTTCAG 1501
Db 222523 CTAAGCTAAGCACTACCTCCGCTTCCACAGGCGCTTTTCAG 222479

RESULT 12
BC023655 1820 bp mRNA linear PRI 30-JUN-2004
LOCUS
DEFINITION
Homo sapiens pleiomorphic adenoma gene-like 2, mRNA (cDNA clone
MGC:29597 IMAGE:4764127), complete cds.
ACCESSION
BC023655
VERSION
BC023655.2 GI:33988277
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1820)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shevchenko, Y.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Schnur, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1820)
Strausberg, R.
Direct Submission
Submitted (05-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 20, 2003 this sequence version replaced gi:23271119.
Contact: MGC help desk
Email: cgapsb-remail.nih.gov
Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nih.gov
Akter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

```
QY 458 AAGGTATGTTTGAACACTTTTAAAGCACGGGAGTGTCTTCTGAGCACCTTAAATCTCAT 517
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 678 AAGGTGTGCTGCAGACCTTTGAGAGTACCAGGCCCTCTAGAGCACCTGAAGGCCAC 737
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 GCAGCAAGTCTGCTGGTGGGTTAAAGAAAAAAGCACAGTGGGAAATTTGTATGTCG 577
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 738 TCACGCGGGTAGCAGCGGTCGCAAGGAGAAAGACACCCCTGTGACCACTGCGACGG 797
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 578 CGGTTCTACACCCGAAAGGATGCGGAGACATGTTGTCACACTGGAAGAAGAC 637
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 798 CGGTTCTACTCGTAAGGATGTACGGCGGACCTAGTGTGTGACACAGCGCGTAGGAC 857
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 TTCCTCTGTAGTATGTGCAAGAGATTTGGGCGAAAGGATCACTGACTCGACATATG 697
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 858 TTCCTGTGCCAGTACTGTGCCAGCGGTTTGGCCGTAAAGACCACTGAGCGTATGTC 917
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 698 AAGAGAGTCAAAATCAAGAGCTTCTGAAGTCAAAACAGAACAGTGGATTTCTTGAC 757
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 918 AAGAGAGCACTCGCAGAGCTGCTCAAGATCAAGACAGAGCGCGTGACATGTTAGGC 977
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 CCAATTACCTGCAATGTGTGTGCTATAAAGACGAGCTCTTCCGCTGATGCTTCA 817
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 978 CTACTCAGCTGCAGCTCCACAGTCACTGTGAAGAGAGCTGAGCCCTGTGCTGCAATG 1037
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 CTTTCCAGTGA---ACTGTTATCAAGCCATTCACAAACACTTTGACGTTAAACCTTAC 874
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1038 GCCTCTCGGAGCGTAATGGGACCAAGGCTTCCCTGCGATGTTGCCCATGGCATGTAT 1097
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 AACACTCCATTTCAGTCCATCAGAGCTCGGAGTCTGCCCA---CCAAATGATCAACT 931
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1098 GGTGCCACATCCCTACCATGCCAGACGCGGCTGCCACACTCCCTGGTGCACAAACG 1157
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 TTACTTTGGGAATGACATGCCCAATAGATATGACACTGTTCATCCCTCTCACACCTT 991
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1158 CTGCCATGGTATGAGTACCTCTGGAATCCCTACCTATCTCTTCCCGAGCTCAGCTC 1217
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 992 TCITTCAAATATCGGTTCACTTCACTCATPATGCAATTTCTATTCCTGAAAAAAGACAG 1051
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1218 CCTCCAAATATACCAGCTTGGATCTACCTCATACTTGC-----CCGACAAA 1262
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1052 CCATTAAGGGGAAATTTGAGGTTACTGATGAGTTACAGGTGGGCTGCGCTCTTCA 1311
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1263 TTGCCCAAGTGGAGTGGATAGTTTCTTGGCGGAGCTTCTGGAAGCCTGTCTCTCA 1332
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1112 TCCCAAGATTTCAAGCATCGTCATCATCTAAAGCTAGGTTGGATCCTCAGATTTGGTCC 1371
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1323 TCCGCTGAACCCAGCCCGCTCA-----CCTCAGCGCGGCGCA 1361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1172 CTAGATGATGGTGCAGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCC 1231
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1362 GCTGGGCGCTCTCCTAGATGAAGCACTGCTTGCACAGAGCCCGCAACCTCTCTGAGCC 1421
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1232 CTAAACACACACGATTTGATTTTCTCAGTTCTGTTAAATTTTCACTTTAAATGTCCT 1291
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1422 CTCTGCGCTGCTAATGTGACCTTCTCCCACTACTGGGCTTTCTTCCACTCAACCTGCC 1481
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1292 CCTATAAA---TCCTCTATCAGTGGGAGCTTGGATGAGCTATTCCAGGAGAGCA 1348
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1482 CCGTGTAAACCCACTTGGGCGCACAGAGGCTGCTGATGGGCTACTCCAGGCTGAGGCA 1541
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1349 CATCTCTGTGTTCCAGCTTCCCAACAAACAGGATCTTCAGGATCCTGCAAAACACT 1408
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1542 CAGCCCTGCTTACCACTTTGAAGCTCAGCTCAAGATTTCCCGAGCTGGGAGCA 1601
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1409 ATAGGCTTGGGCTCTGCACTCACTGTGAGAGCTTTTCCAGAGATTTAAGCAAGT 1468
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1602 CTGAATTTGGGCTCTGCACTCTTGGCTCTGCTTCAAGCTGCTGGCTGAGT---AGC 1658
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1469 ACCACCTCCCAAGTTTCCATCAAGCTTTTCAG 1501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1659 ACCACCTGCTTCTTCCATCAAGCATTTCCAG 1691
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 13
CQ834146

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

ORIGIN

Query Match

Best Local Similarity

Matches 823; Conservative

26.0%; Score 389.8; DB 6; Length 2335;

58.2%; Pred. No. 5e-103;

Mismatches 542; Indels 48; Gaps 6;

QY 98 AACTTCTCTGCAACTGTGTGACAGGCTTTTAAAGTGTGAGAAATTAAGTTTCA 157

Db 125 AAGTGCATGTGAAATTTTCGGAACACCTTTCTCAAATGGGAGAGCTGAGGCTTCA 184

QY 158 TCCTACTCTCACACAGGAGAGGCTTCAAGTGCATACAAAGACTGCACCAAGGCC 217

Db 185 AGCTCCGCAACAGAGAGAGACCATATAGTGCCTCAGCTGCACTGTGGCAAGCT 244

QY 218 TTTGTTTCTAAGTACAAATTTACAAAGCAGATGGTACTCATTTCTCTGAGAAAAACCCAC 277

Db 245 TTTGTTTCTAAGTACAAATTTACAAAGCAGATGGTACTCATTTCTCTGAGAAAAACCCAC 304

QY 278 AAGTGTAAATTTTGTGAGAAATTTTCCCGGAAAGATCATCTGAGAAATCACTTCAAT 337

Db 305 CAGTGTATGTATGTATGATAGATTTTCCCGGAAAGACCATCTGGGAAACCATCTGCAG 364

QY 338 ACACAGACCCCTAACAAAGAGAGCTTTAAGTGCAGAAATTTGCGAAGACTGCAGAACTACATACC 397

Db 365 ACCATGATCTTACAAAGAGGCTTCCCTCCTGCTCTGAGTGGGTAAAGATTAACAATACG 424

QY 398 AAGCTTGGATTTAAAGCTCACTTGGCTTTGATGCGCAACAAAGTGGTGGTCACTTGT 457

Db 425 AAGCTGGGTACCGCGCCACCTGGCCATGATGCTGCCAGAGCGGTGACCTCAGCTGC 484

QY 458 AAGGTATGTTTCAAACTTTTGAAGACAGGAGTGTCTTCTGGAGCACTTAAATCTCAT 517

Db 485 AAGGTGTGCTGACAGACCTTTTGAGAGTACCCAGGCGCTGTAGAGCACTGAGGCGCCAC 544

QY 518 GCAGGCAAGTCTGCTGGTGGGCTTAAAGAAAAAAGACCACTGAGCGAACTTGTATCGC 577

Db 545 TCACCGCGGTAGCAGGCGGTGCCAAGGAGAAAGACCCCTCTGTGACCACTGCGACCGG 604

CQ834146 2335 bp DNA linear PAT 29-JUL-2004

Sequence 17 from Patent WO2004058805.

CQ834146

CQ834146.1 GI:50833683

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Matsuda, A. and Yoneta, S.

T cell activating gene

Patent: WO 2004058805-A 17 15-JUL-2004;

Asahi Kasei Pharma Corporation (JP)

Location/Qualifiers

1..2335

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

11..1501

/note="unnamed protein product"

/codon_start=1

/protein_id="CAH05288.1"

/db_xref="GI:50833684"

/translation="MTTFTSVPPWIQDAKQEEVEVGMKLVPRPRGREAESOVKQCEI

SGTSPFNGEKLRLPHLQTHDNKEALHCEGKNVTKLGYRHLHMAHSAAGDLSCKV

CLDTFSTQALLLEHLKAHSRRVAGAKKHPDCHDRFRYTRKDRVRLHVLVGRKD

FLQCYCAQRGRKDLTRHVKSHSDELLKIETPEVDMGLLSCSTVSVKBELSPVL

CMASRDVMTGKAPFPLMGVGMGIAHITMPESTGVPSLHVNTPMGSVPLESSPIS

PAQLPPKYLQSTSYLPDLKPKVEVDSFLAELPGLSLSSAEPQSPQAAAAALLD

EALLAKSPANLSEALCAANVDFSHLLGFLPLNLPNPPGATGSLVMGYSQAEQPLL

TTLQAQPDSPGAGPLNFGPLHSLPVPVTSGLSSITLPRHQAFQ"

QY	578	CGGTTCTACACCCGAAGGATGTCGGAGACACATGTTGGTGCACACTGGAGAAAGAC	637
Db	605	CGGTTCTATACCTCGTAAGGATGTCGGCGGACACCTAGTGGTGCACAGCGCGTAAGAC	664
QY	638	TTCTCTGTCTAGTATTGTGCACAGAGATTTGGGGGAAAGGATCCTGACTCGACATAG	697
Db	665	TTCTCTGCCAGTACTGTGCCAGCGGTTTGGCGGTAGGACCACTGACGCGTCATGTC	724
QY	698	AAGAGAGTCAATCAAGAGCTTCTGAAGCTCAAAACAGAACCGAGTTCCTTAC	757
Db	725	AAGAGAGCCACTGCGAGAGCTGCTCAAGATCAAGACAGAGCCGTCGACATGTTAGGC	784
QY	758	CCATTACTGGAATGTCTGTGCTATAAAGACGAGCTTCCTCCGCTGATGTCCTTA	817
Db	785	CTACTCAGCTGCAGCTCCACAGTCACTGTGAAGGAGAGCTGAGCCCTGTCTGTGCAATG	844
QY	818	CCTTCCAGTGA---ACTGTTATCAAAAGCCATTCAAAAACATTTGCAGTTAAACCTCTAC	874
Db	845	GCCTCTCGGACGTAATGGGACCAAGGCCTTCCCTGGCATGTTGCCATGGGCATGTAT	904
QY	875	AACACTCCATTTCACTGATGAGAGCTCGGATCTGCCCA---CCAAATGATCAACT	931
Db	905	GGTGCCCAACATCCCTACCATGCCAGCAGCGGCTGCCACACTCCCTGGTGACACAGC	964
QY	932	TTACTTTGGGAATGACATGCCCCAATAGATATGGACACTGTTTCATCCCTCTCACCACTT	991
Db	965	CTGCCCATGGTATGAGCTACCTCTGGAACTCTACCTATCTCTTCCCAGCTCAGCTC	1024
QY	992	TCTTTCAAAATATCCGTTCACTTCTACCTCATATGCAATTTCTATCTCTGAAAAGAACAG	1051
Db	1025	CCTCCAAATACAGCTTGAATCTACTCTACTTGC-----CCGACAAA	1069
QY	1052	CCATTAAAGGGGAAATGAGATTACCTGATGAGATTACAGGTGGGTGGCTCTTCA	1111
Db	1070	TTGCCCAAAGTGAGGTGGATGTTTCTGGCGAGCTTCTTGAAGCCCTGTCTCTCTCA	1129
QY	1112	TCCCAAGATTCTCAAGCATGCTCATCATCTAAGCTAGGTGGATCTCAGATTGGGTGC	1171
Db	1130	TCCGTGAACCCAGCCCGCCTCA-----CCTCAGCGCGGCA	1168
QY	1172	CTAGATGATGGTGCAGGAGACCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCC	1231
Db	1169	GCTGGCGCCCTCCTAGATGAAGCACTGCTTGGCAAGAGCCCGCAACCTCTCTGAGGC	1228
QY	1232	CTAAACACACAGATGGATTTTCTCAGTTGTTTAAATTTTCACTTTAAATGGTCT	1291
Db	1229	CTCTGGCTGCTAATGTGACTTCTCCCACTACTGGGCTTCTTCCACTCAACCTGCC	1288
QY	1292	CCCTATAA---TCTCTATCAGTGGGAGCTTGGATGAGCTATTTCCAGGAAAGCA	1348
Db	1289	CCGTGTAACCCACTGGGGGCAAGGAGGCTGTCTATGGCTACTCCAGGCTGAGGCA	1348
QY	1349	CATTCTCTCTTTCCAGCTCCCAACAAACAGGATCTTCAAGATCTTGCAAACT	1408
Db	1349	CAGCCCTGTCTACCACTTTGCAAGCTCAGCCTCAAGATTTCCAGGAGCTGGGGACCA	1408
QY	1409	ATAGGGCTTGGTCTGTGACTACTGTACAGCTTTCACAGCAGTTTAAAGCACAAGT	1468
Db	1409	CTGAACCTTTGGGCTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1465
QY	1469	ACGACCTTCCACGTTTCCATCAAGCTTTTTCAG	1501
Db	1466	ACCACTGCTCTGCTTTTCCATCAAGCATTCAG	1498
RESULT 14			
AK026936			
LOCUS			
DEFINITION			
Homo sapiens cDNA: FLJ23283 f1s, clone HEP08729, highly similar to			
AF006005 Homo sapiens zinc finger protein PLAGL2 mRNA.			
ACCESSION			
AK026936			
VERSION			
AK026936.1 GI:10439912			

KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
		Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1	
AUTHORS		Kawabata, A., Hiki, J., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.	
		NEDO human cDNA sequencing project	
TITLE		Unpublished	
JOURNAL		2 (bases 1 to 2429)	
REFERENCE		Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.	
AUTHORS		Direct Submission	
TITLE		Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: f1cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)	
JOURNAL		NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).	
COMMENT			
FEATURES		Location/Qualifiers	
SOURCE		1..2429	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="HEP08729"	
		/cell_line="HepG2"	
		/clone_lib="hepatoma"	
		/note="cloning vector pME18SPL3"	
misc_feature		1..2429	
		/note="highly similar to AF006005 Homo sapiens zinc finger protein PLAGL2 mRNA"	
ORIGIN			
Query Match		26.0%; Score 389.8; DB 9; Length 2429;	
Best Local Similarity		58.2%; Pred. No. 5e-103;	
Matches 823; Conservative		0; Mismatches 542; Indels 48; Gaps 6;	
QY	98	AACTTTCTCTGCACTGTGTGACAAAGGCGCTTTAAACAGTGTTCAGAAATTAAGGTTTCA	157
Db	199	AACTTTCTCTGCACTGTGTGACAAAGGCGCTTTAAACAGTGTTCAGAAATTAAGGTTTCA	258
QY	158	TCTTCTCTCAGAGAGAGAGGCGCTTACAAAGGCGCTTTAAACAGTGTTCAGAAATTAAGGTTTCA	217
Db	259	AGCCTCCGCAACAGAGAGAGGCGCTTACAAAGGCGCTTTAAACAGTGTTCAGAAATTAAGGTTTCA	318
QY	218	TTTGTCTTCTAAGTACAAATTAACAAAGGCGCTTTAAACAGTGTTCAGAAATTAAGGTTTCA	277
Db	319	TTTGTCTTCTAAGTACAAATTAACAAAGGCGCTTTAAACAGTGTTCAGAAATTAAGGTTTCA	378
QY	278	AAAGTGAATTAATTTGTGAGAAATTTTCAACCGGAAAGATCATCTGAAGATTCACCTCCAT	337
Db	379	CAGTGTATGATGTGTGATAAGTGTTCACCGGAAAGATCATCTGAAGATTCACCTCCAT	438
QY	338	ACACAGACCTTACAAAGAGAGCGTTTAAAGTGAAGATTCAGAAATTAAGGTTTCA	397
Db	439	ACCATGATCTTAAAGAGAGCGTTTAAAGTGAAGATTCAGAAATTAAGGTTTCA	498
QY	398	AAAGTGTGATTTAAACGTCCTTGGCTTGCATCCGCAACAGTGTTCAGAAATTAAGGTTTCA	457
Db	499	AAAGTGTGATTTAAACGTCCTTGGCTTGCATCCGCAACAGTGTTCAGAAATTAAGGTTTCA	558
QY	458	AAAGTGTGATTTAAACGTCCTTGGCTTGCATCCGCAACAGTGTTCAGAAATTAAGGTTTCA	517
Db	559	AAAGTGTGATTTAAACGTCCTTGGCTTGCATCCGCAACAGTGTTCAGAAATTAAGGTTTCA	618

QY	518	GAAGCAAGTCGTCTGTGGGGTTTAAAGAAAAAGACACAGTGGGAACATTTGTGATCGC	577
DB	619	TCAAGCGGGTAGCAGCGCGTCCAAAGAGAGAGACCCCTGTGACCACTGCGACCGG	678
QY	578	CGGTCTACACCCGAAAGGATGTCCGGAGACAATGGTGTGCACACTGGAAGAAGGAC	637
DB	679	CGGTTCTATATCTGTAAGGATGTACGGCGGACCTAGTGGTGACACAGCGCGTAAGNC	738
QY	638	TTCTCTGTCACTATTGTGCACAGAGATTTGGCGGAAGGATCACCTGACTCGACATATG	697
DB	739	TTCTGTGCCAGTACTGTGCTAGCGTTTGGCGGTAAGGACCACTGACGGGTCAATGTC	798
QY	698	AAGAGAGTCACAATCAAGAGCTTCTGAAGGTCAAAACAGAACCAAGTGGATTTCTTGAC	757
DB	799	AAGAAGACCACTCGCAGGAGCTGCTCAAGATCAAGACAGAGCCGTGACATGTTAGGC	858
QY	758	CCATTACCTGCAATGTGTCTGCGCTATAAAGACAGACTCTCTTCGGTGTATGCTTTA	817
DB	859	CTACTCAGCTGCAGTCCACAGTCAAGTGTGAAGGAGAGCTGAGCGCTGTGTGTGCATG	918
QY	818	CTTCCAGTGA---ACTGTTATCAAGCCATTCACAAACACTTTCAGTTAAACTCTTAC	874
DB	919	GCCTCTCGGAGCAATAATGGGACCAAGCGCTTCCCTGCAATGTGCCCATGGCATGTAT	978
QY	875	AACACTCCATTTCAGTCCCATGCAGAGCTCGGGAATCTGCCA---CCAAATGATCACAACT	931
DB	979	GGTGCCCAATCTCCATCCATGCCAGCAGCGGTGCCACACTCCCTGGTGACAAACAG	1038
QY	932	TTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCATCCCTCTCACACTT	991
DB	1039	CTGCCCATGGGTATGAGCTACCTCTGGAATCTCTCACCTATCTCTTCCCAGCTCAGCTC	1098
QY	992	TCCTTCAATATCGGTTTCAGTTTACCTCATATGCATTTCTATNTCTGAAAAAGAACAG	1051
DB	1099	CTCCAAATACCAAGCTTGGATCTACCTCATACTTGC-----CCGACAAA	1143
QY	1052	CCATTTAAGGGGGAATTTGAGAGTTTACTGTATGAGTTACAGGTGGGTGCCCTCTTCA	1111
DB	1144	TTGCCCAAGTGGAGTGATGTTTCTTGGCGAGCTTCTGGAAGCCGTGCTCCCTCTCA	1203
QY	1112	TCCCAAGATTCGAAGCATCTCATCATCTAAAGCTAGGTTGGATCCCTCAGATTGGTGTC	1171
DB	1204	TCCGCTGAACCCAGCCGCCCTCA-----CCTCAGCCGGCGCA	1242
QY	1172	CTAGATGATGTGCAGGAGACCTCTCCCTATCMAAAGCTCTATCTCOATCAGTGACCGC	1231
DB	1243	GCTGGGCGCTCTATAGTAGAAGCACTGCTTGCCTCAAGAGCCCGCCAACTCTCTGAGGCC	1302
QY	1232	CTAACAACACCAAGATTGGATTTTCTTCAGTTGTTTAAATTTTCACTTTAAATGTCCT	1291
DB	1303	CTCTGCGCTGCTAATGTGGACTTCTCCCACTACTGGGCTTCTTCCACTCAACTCGGCC	1362
QY	1292	CCCTATAA---TCTCTATCAGTGGGAGCGTTTGGAAATGAGCTATTTCCAGGAAGAACA	1348
DB	1363	CCGTGTAAACCACTGGGGCCACAGGAGGCTGCTCATGGGCTACTCCCAAGCTGAGGCA	1422
QY	1349	CAATCTTCTGTTTCCCAGCTCCCCACACAAACAGGATCTTCAGGATCTCTGCAAACT	1408
DB	1423	CAGCCCTGCTTACCATTGTGAAGCTCAGCTTCAAGATTTCCCAAGAGCTGGGGGACA	1482
QY	1409	ATAGGGTTGGGTCTGTGCACTCACTGTCAAGCAGTCTTCCACGAGCTTTTAAAGCAAGT	1468
DB	1483	CTGAACCTTTGGGCTGTGCACTCCTTGCTCTCTTCACTCTGCTGCTGCTGCTGCTGCT	1539
QY	1469	ACCAACCTCCCAAGTTTCCATCAAGCTTTTCAG	1501
DB	1540	ACCAACCTGCTGCTTTCATCAAGCAATTCAG	1572

RESULT 15
AF006005
LOCUS

DEFINITION	Homo sapiens zinc finger protein PLAGL2 (PLAGL2) mRNA, complete cds.
ACCESSION	AF006005
VERSION	AF006005.1
KEYWORDS	GI:3513455
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 4051)
TITLE	Kas,K., Voz,M.L., Hensen,K., Meyen,E. and Van de Ven,W.J.
FEATURES	Transcriptional activation capacity of the novel PLAG family of zinc finger proteins
JOURNAL	J. Biol. Chem. 273 (36), 23026-23032 (1998)
MEDLINE	98389728
PUBMED	9722527
REFERENCE	2 (bases 1 to 4051)
AUTHORS	Kas,K., Hensen,K., Meyen,E., Voz,M.L. and Van de Ven,W.J.M.
TITLE	Direct Submission
JOURNAL	Submitted (29-MAY-1997) Laboratory for Molecular Oncology, Center of Human Genetics, Herestraat 49, Leuven 3000, Belgium
FEATURES	Location/Qualifiers
source	1. 4051
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/chromosome="20"
	/map="20q11"
	1. 4051
gene	/gene="PLAGL2"
CDS	203..1693
	/gene="PLAGL2"
	/note="C2H2-type zinc finger protein; PLAG-like 2"
	/codon_start=1
	/product="zinc finger protein PLAGL2"
	/protein_id="AAC34252.1"
	/db_xref="GI:3513455"
	/translation="WTTFFTSVPPWIDAKQEEVGVKLVPRPRGAEASOVKQCCG SGTFFSNGEKLRHLSLPQPCQSCPLQHCQKAFASKYKLYRHMAHSAQKPHOCM CKDMFHRKHLRHLQPDHNKEALHCSECKNYNTKLYRRHLAMHAAASGDLSCRY CLQTFSTQALLRHLKAHSRRVAGAKKPCDCHCDRFFYTRKDVRRHLVHTGRK FLCQYCAQREGRDHLTRHVKKSHSQBLIKITFEDVMLGLLSCSTVSVKBEELSPV CMASRDVMGFKAPGMLPMGMYGNAHITMFSTGVPHSLVHNTLPMGMSYPLESSPIS PAQLPKYOLGTSYLPDLKPKVEVDSFLAELPGSLSLSSAEPQAFQFPAAAAALIL EALLAKSPNLSSGALCAANVDVFLSHLLGFPLNLPCCNPPGATGIVMGVSOAEAQPI TILQASQSPGAGGVDFPHLSLPVFTSGSLSTTLPRFHOAFQ"

ORIGIN

Query Match	26.0%;	Score 389.8;	DB 9;	Length 4051;
Best Local Similarity	58.2%;	Pred. No. 5e-103;		
Matches 823;	Conservative 0;	Mismatches 542;	Indels 48;	Gaps 6;
Qy	98	AAC TTTCTCTTGC CAACAGTGTGCAAGGCGCTTTAACACAGTGTGAGAAATTAAGGTTCCAC	157	
Db	317	AAGTGC CAATGTGAAATTTTCGGGAACACCTTTCTCAAATGGGGAAGCTGAGGCTTCAC	376	
Qy	158	TCCTACTCTCACAGAGAGAGCGCCCTCAAGTGCATACAAAGAGACTGCACCAAGGCC	217	
Db	377	AGCCTCCGCAACAGAGAGAGACCATATAGTGCCTCAGCTGCACCTGTGCGNAGGCT	436	
Qy	218	TTTGT TTTCTAAGTACAAATTAACAAGGCACATGGCTACTCATTTCTCTGAGAAAACCCAC	277	
Db	437	TTTGTCTTCCAAATACAAGCTGTATAGGCACATGGCACCCTCAGCCCGAAGAACCCAC	496	
Qy	278	AAGTGTAAATATTGTGAGAAAATGTTTCCACCGAAAGATCATCTGAAGATACCTCCAT	337	
Db	497	CAGTGTATGTACTGTGATGAAGATGTTTACCGCAAGGACCATCTCGGGAACCATCTGCAG	556	
Qy	338	ACACAGCCCTTAACAAGAGAGGTTTAAAGTGCAGAAATGTGGCAAGAACTACAATACC	397	
Db	557	ACCATGATCCTTAACAAGAGGCGCTCCACTGCTCTGAGTGGGTAAGAAATTACAATACG	616	
Qy	398	AAGCTTGGATTTAAACGTCAC TTGCGCTTGCATGCGCGAACAAAGTGGTGAACCTCACTGT	457	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2004, 09:58:19 ; Search time 786 Seconds
(without alignments)
10024.658 Million cell updates/sec

Title: US-09-242-772-116_COPY_480_1980
Perfect score: 1501
Sequence: 1 gatggcactgtcattccg.....gtttcatcaagctttttcag 1501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23sep04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002as:*

7: Geneseq2002bs:*

8: Geneseq2003as:*

9: Geneseq2003bs:*

10: Geneseq2003cs:*

11: Geneseq2003ds:*

12: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1501	100.0	7313	2	AAV29268
2	351.6	23.4	2334	2	AAV18481
3	351.6	23.4	2828	6	ABN86525
4	351.6	23.4	3215	11	ADM02534
5	350	23.3	2815	6	ABL61932
6	350	23.3	4632	6	ABN86524
7	348.4	23.2	3991	12	ADQ24242
8	301.8	20.1	5028	6	ABN86508
9	301.8	20.1	5028	10	ABT41834
10	294.4	19.6	2790	2	AAV18480
11	294.4	19.6	3975	6	ABN86526
12	294.4	19.6	3975	12	AD126094
13	277.6	18.5	2561	2	AAV29269
14	277.6	18.5	2561	6	ABT11028
15	276	18.4	2738	4	AAZ33549
16	267	17.8	475	4	AAI10993
17	267	17.8	475	4	ABAS2644
18	267	17.8	475	4	AAI32253
19	267	17.8	475	4	ABAS2644
20	267	17.8	475	4	ABAS2644
21	267	17.8	475	4	AAK26360
22	267	17.8	475	4	AAK00905

C	22	267	17.8	475	4	ABS25951	Human liv
C	23	267	17.8	475	5	AAI00914	Probe #90
C	24	267	17.8	475	6	ABS00948	Human gen
C	25	267	17.8	1083	12	ADP28687	ADP28687 Human sec
C	26	131.4	8.8	230	4	AAI20227	Probe #10
C	27	131.4	8.8	230	4	ABA65256	ABA65256 Human foe
C	28	131.4	8.8	230	4	AAI45427	AAI45427 Probe #14
C	29	131.4	8.8	230	4	ABA32357	ABA32357 Probe #10
C	30	131.4	8.8	230	4	AAK39414	AAK39414 Human bon
C	31	131.4	8.8	230	4	AAK13673	AAK13673 Human bra
C	32	131.4	8.8	230	4	ABS39002	ABS39002 Human liv
C	33	131.4	8.8	230	5	AAI05931	AAI05931 Probe #59
C	34	131.4	8.8	230	6	ABS13501	ABS13501 Human gen
C	35	126.8	8.4	324	6	ABL77615	ABL77615 Human ova
C	36	84.8	5.6	445	9	ACH41060	ACH41060 Human foe
C	37	72	4.8	72	4	AAI31082	AAI31082 DNA encod
C	38	72	4.8	72	8	ACC41633	ACC41633 Human zin
C	39	65.8	4.4	2540	10	ADC37636	ADC37636 Human nuc
C	40	65.8	4.4	2864	4	AAH14677	AAH14677 Human cDN
C	41	64.2	4.3	1449	12	ADO00620	ADO00620 Novel hum
C	42	64.2	4.3	1449	12	ADN99051	ADN99051 Novel hum
C	43	64	4.3	580	12	ACH92023	ACH92023 Human gen
C	44	63.4	4.2	1995	12	ADN98725	ADN98725 Novel hum
C	45	63.4	4.2	1995	12	ADO00294	ADO00294 Novel hum

ALIGNMENTS

RESULT 1
AAV29268
ID AAV29268 standard; cDNA; 7313 BP.
XX
AC AAV29268;
XX
DT 21-AUG-1998 (first entry)
XX
DE Nucleotide sequence of human PLAG1.
XX
KW Human PLAG1 gene; PLAG1; tumorigenesis gene; T-gene; PLAG2; CTNNB1;
KW antibody; benign tumour; malignant tumour; leukaemia; cancer;
KW inhibition; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 481..1983
FT /*tag=a
FT /product= "PLAG1 protein"
FT /transl_except= (pos: 1603..1605, aa: Gly)
FT /transl_except= (pos: 1861..1863, aa: Gly)
XX
XX EP825198-A1.
XX
XX
PD 25-FEB-1998.
XX
XX
PP 17-JAN-1997; 97EP-00200130.
XX
XX 22-AUG-1996; 96EP-00202339.
XX
XX (KULB-) KU LEUVEN RES & DEV.
XX (UYGO-) UNIV GOETEBORGS HOLDINGBOLAGET AB.
XX Van De Ven WJM, Stenman KGD, Kas KP, Voz ML;
XX WPI; 1998-132252/13.
XX P-PSDB; AAW37948.
XX
XX New tumorigenesis T-genes and proteins - useful for, e.g. preparing
XX antibodies for clinically diagnosing cells having non-physiological
XX proliferative capacity such as lipoblastomas.
XX
PS Claim 2; Fig 4; 7lpp; English.

XX	Sequence	7313 BP; 2287 A; 1327 C; 1356 G; 2342 T; 0 U; 1 Other;
SY	Query Match	100.0%; Score 1501; DB 2; Length 7313;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 1501; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GATGGCCACTGTCAATTCCTGGTGATTTGTTCAGAACTAAGAGATACCCAGAAAGTCCTTC 60
DB	480	GATGGCCACTGTCAATTCCTGGTGATTTGTTCAGAACTAAGAGATACCCAGAAAGTCCTTC 539
QY	61	AGGAAACGTAAAGCGGTGTAACACCAACCAACCAAGAAAACTTTCCTTGCACCTGTGTGA 120
DB	540	AGGAAACGTAAAGCGGTGTAACACCAACCAACCAAGAAAACTTTCCTTGCACCTGTGTGA 599
QY	121	CAAGGCGCTTTAAACAGTGTTCAGAAATTTAAAGGTTCACTCTCTCACACAGGAGAG 180
DB	600	CAAGGCGCTTTAAACAGTGTTCAGAAATTTAAAGGTTCACTCTCTCACACAGGAGAG 659
QY	181	GCCCTAACAGTGATCAACAAAGACTGCACCAAGGCTTTGTTTCTAAGTACAATTACA 240
DB	660	GCCCTAACAGTGATCAACAAAGACTGCACCAAGGCTTTGTTTCTAAGTACAATTACA 719
QY	241	AAGSCATGCTACTCATTTCTCTGAGAAACCCCAAGTGTATAATTATTGTGAAAAAT 300
DB	720	AAGSCATGCTACTCATTTCTCTGAGAAACCCCAAGTGTATAATTATTGTGAAAAAT 779
QY	301	GTTTCAACGGAAAGATCATCTGAAGATCACTCCATACACAGCAACCTTAAACAAGAGAC 360
DB	780	GTTTCAACGGAAAGATCATCTGAAGATCACTCCATACACAGCAACCTTAAACAAGAGAC 839
QY	361	GTTTAAAGTGCAGAAATCTGGCAAGAACTTCAATACCAAGCTTGAATTTAAACGTCATT 420
DB	840	GTTTAAAGTGCAGAAATCTGGCAAGAACTTCAATACCAAGCTTGAATTTAAACGTCATT 899
QY	421	GCCCTTGATGCGCGAAACAAAGTGGTGAACCTACCTGTAAAGTATGTTTGCACAACTTTGA 480
DB	900	GCCCTTGATGCGCGAAACAAAGTGGTGAACCTACCTGTAAAGTATGTTTGCACAACTTTGA 959
QY	481	AAGCAGGGAGTGCCTTCTGGACACCTTAAATCTCATGACGCAAGTCGTCGTGGGGT 540
DB	960	AAGCAGGGAGTGCCTTCTGGACACCTTAAATCTCATGACGCAAGTCGTCGTGGGGT 1019
QY	541	TAAAGAAAAAGACCAAGTCGGAACATTTGTATGCCCGGTTCTACACCGAAAGATGT 600
DB	1020	TAAAGAAAAAGACCAAGTCGGAACATTTGTATGCCCGGTTCTACACCGAAAGATGT 1079
QY	601	CCGGAGACATGTTGGTGTCAACACTGGAAGAAAGACTTCTCTGTCAAGTATTTGTGCACA 660
DB	1080	CCGGAGACATGTTGGTGTCAACACTGGAAGAAAGACTTCTCTGTCAAGTATTTGTGCACA 1139
QY	661	GAGATTGGCGGAAAGATCAACCTGACTGCACATATGAAGAAGATCAATCAAGAGCT 720
DB	1140	GAGATTGGCGGAAAGATCAACCTGACTGCACATATGAAGAAGATCAATCAAGAGCT 1199
QY	721	TCTGAAGGTCAAAACAGAACAGTGGATTTCTTGCACCAATTTACCTGCAATGTCTGT 780
DB	1200	TCTGAAGGTCAAAACAGAACAGTGGATTTCTTGCACCAATTTACCTGCAATGTCTGT 1259

RESULT 2

QY	64	GA	AACGTAAGCGTGGTGAACCAAAACCAAGAAAAAACTTCTTGCCAACTCTGTGACAA	123
Db	1269	GA	AAGGAAAGGTTGTGAGAAGCAAGCCATGGCCAGTTCCCTGCCAGTTATGTGGCAA	1328
QY	124	GG	CTTTAAACAGTGTGTGAGAAAATAAAGGTTCACTCTACTCTCAACAGGAGAGAGGCC	183
Db	1329	GA	CTTCTCAACCTGGAGAAAGTTACGATTACAATTTATCCCACTCCAGGAGCGGC	1388
QY	184	CT	CAAGTGCATACAACAAGACTGCACCAAGGCTTTGTTCTTAAGTACAAATTACAAG	243
Db	1389	GT	ACAAGTGTGCAGCCTGACTGTGCAAGGCTTTGTTTCAGATATAAAATTGATAG	1448
QY	244	GCA	CATGGCTACTCAITTCCTCTGAGAAAACCCACAAGTGTAATTTATGTCAGAAAAATGTT	303
Db	1449	GC	ATATGGCTACCCATTCTCCCCAGAAATCTCACAGTGTCTCACTGTGAGAAAGCTTT	1508
QY	304	TC	CGGAAAGATCATCTGAAGAATACCTCCATACACAGACCTTAACAAAGAGAGCTTT	363
Db	1509	CA	ACCGGAAGACCACTGTAAAAACCACTCCAGACCCACAGCCCAACAAATGGCCTT	1568
QY	364	TA	AGTGCGAAGAATGTGSCAAGAACTACAATAACAAAGTTGGATTAAAGTCACTTGGC	423
Db	1569	TGG	TGTGAGAGTGTGGGAAGAAGTACAACACCATGTGGGCTATAGAAGGCACCTGGC	1628
QY	424	CT	TCATGCGCGCAACAAGTGTGACCTCAACCTGTAAAGTATGTTGCAAACTTTTGAAG	483
Db	1629	CT	CCATGCGGCGACAGCTGGGGACCTCACTGTGGGTCTGTGCCCTGAGCTAGGAG	1688
QY	484	CA	CGGAGTGTCTTCGGAGCACCTTAAATCTCATGC---AGGCAAGTCTGTCTGTGGGT	540
Db	1689	CA	CCGAGTGCTACTTGGACCACTCTAAAGCCATGCGGAAGAGGCCCTTAGCGGAAC	1748
QY	541	TA	AGAAAAAAGCACCAAGTGCACATTTGTATCGCGGTTCTACCCGAAAGGATGT	600
Db	1749	CA	AGGAAGAAGACCAAGTGCACCACTGTGAAAGATGCTTCTACACCGGAAGGATGT	1808
QY	601	CC	GGAGACACATGTTGGTGCACACTTGGGAAGAAGACTTCTCTGTCTAGTATTGTGACA	660
Db	1809	CG	AGCCACCTTGGTGTCCACACAGGATGCAAGGACTTCTGTGCCAGTCTGTGCCCCA	1868
QY	661	GAG	ATTGGGCAAGGATCACTGACTCGACATATGAAGAAGTGTCACAATCAAGAGCT	720
Db	1869	GAG	ATTGGGCGCAAGGATCACTCACCCGGCATTACCAAGAGACCCACTCACAGGACT	1928
QY	721	TC	TGAAGGTCAAAAACAGAAACCAAGTGATTTCCTTGAACCAATT	762
QY	1990	GT	CTCAAGCACTTTCGACACCGGAGACTTCTTGAGCAGCTT	1970

RESULT 4
ADM02534
IN ADM02534 standard: CDNA: 3215 BP.

AC ADM02534;

XX
20-MAY-2004 (first entry)

XX Human cDNA of the invention SEQ ID NO:1219.

XX
pharmaceutical.
diagnostic marker:
pharmacovigilance

XX
XX

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

XX

PD
XX
Z4-SEP-2003.

12-APR-2002; 2002EP-00008400.
PF
vv

PR 22-MAR-2002; 2002JP-00137785.

PA

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S, Yamamoto J, Isono Y, Otsuka K, Nagai K, Irie R, Tamechika I, Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y.

WPI: 2003-723558/69.
P-PSDB: ADM04977.

QY 281 TGTAAATTTGAGAAAATGTTTTCACCGGAAGATCATCTGAAGATCACTCCATCAVACA 340
Db |||||
737 TGCACCTATTGTGAAAGACTTTCAACCGGAAGGATCACTGAAGATCACTCCAGACC 796
QY 341 CACGACCTTAACAAGAGACGCTTTAAAGTGCAGAAATGTGGCAAGAACTACAATACCAAG 400
Db |||||
797 CACGATCCCAACAAGATGATCTACGCTTGGAGAGATTTGGCAAGAAATACCAACCATG 856
QY 401 CTGTGATTTAAACGTCACCTGGCCCTTGCATGCCGCAACAAGTGTGACCTCACTGTAAAG 460
Db |||||
857 CTGGCTACAGAGGACATGATGCCCTGCAATTCGGCCAGCAGCGCGATCTCACTGCGGC 916
QY 461 GTATGTTGAAACTTTTGAAGACGAGGAGTGCTTCTGGAGCACTTAAATCTCATGC- 519
Db |||||
917 GTCTGACCCCTGGAGCTGGGAGCACCGAGGTCCTGCTGGACCACTCAAGTCTCACGGC 976
QY 520 --AGCAAGTCTCTGTGGGTAAAGAAAAAAGCAACCAAGTGGCGAAATTTGTGATGC 577
Db |||||
977 GAAGAAAAGGCCACACCGCCCGAGGAGAAACACCACTGGAGCACTGGAGAGA 1036
QY 578 CGGTTCTACACCCGAAGATGTCGGAGACACATGCTGGTGCACACTGGAAGAAGGAC 637
Db |||||
1037 TGCTTCTACACCCGAAGATGTCGCTGCCACCTGGTGCACAGAGATGCAAGGAC 1096
QY 638 TTCCTCTGTCAGTATGTCACAGAGATTTGGCGAAAGGATCACTGACTGCACATATG 697
Db |||||
1097 TTCCTGTGTCAGTCTGCGCCAGAGATTTGGCGGCAAGACCACTCACTGCTCACACC 1156
QY 698 AAGAAGAGTCACAATCAAGAGCTTCTGAAGTCAAAACAGAACCAAGTGGATTTCCCTTGAC 757
Db |||||
1157 AAGAAGACCCATCCACGAGCTGATGCAAGAGAGCCCTGCAAGCAGGAGAAATACCAAGGC 1216
QY 758 CCATTTACCTGCAATGTCT 778
Db |||||
1217 GGTACCACCCATTTGGGCT 1237

RESULT 9
ABT41834
ID ABT41834 standard; DNA; 5028 BP.
XX
AC ABT41834;
XX
DT 26-JUN-2003 (first entry)
XX
DE Toxicity modelling related rat gene SEQ ID No 1536.
XX
KW Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW database; drug screening; toxicity assay; rat; ds.
XX
OS Rattus norvegicus.
XX
PN WO200295000-A2.
XX
PD 28-NOV-2002.
XX
PF 22-MAY-2002; 2002WO-US016173.
XX
PR 22-MAY-2001; 2001US-0292335P.
PR 13-JUN-2001; 2001US-0297523P.
PR 19-JUN-2001; 2001US-0298925P.
PR 10-JUL-2001; 2001US-0303807P.
PR 10-JUL-2001; 2001US-0303808P.
PR 10-JUL-2001; 2001US-0303810P.
PR 28-AUG-2001; 2001US-0315047P.
PR 27-SEP-2001; 2001US-0324928P.
PR 22-OCT-2001; 2001US-0330462P.
PR 01-NOV-2001; 2001US-0330867P.
PR 21-NOV-2001; 2001US-0331805P.
PR 06-DEC-2001; 2001US-0336144P.
PR 19-DEC-2001; 2001US-0340873P.
PR 21-FEB-2002; 2002US-0357842P.
PR 21-FEB-2002; 2002US-0357843P.

21-FEB-2002; 2002US-0357844P.
15-MAR-2002; 2002US-0364134P.
08-APR-2002; 2002US-0370144P.
08-APR-2002; 2002US-0370206P.
08-APR-2002; 2002US-0370247P.
17-APR-2002; 2002US-0372794P.
21-APR-2002; 2002US-0371679P.
XX (GENE-) GENE LOGIC INC.
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
PI WPI; 2003-148464/14.
XX
DR Predicting at least one toxic effect of a compound, useful for toxicity
XX modelling, comprises preparing a gene expression profile of a tissue or
PT cell sample exposed to the compound, and comparing the gene expression
PT profile to a database.
PT
XX Example 4; Page; 446pp; English.
XX The invention relates to a novel method of predicting at least one toxic
CC effect of a compound. The method comprises a gene expression profile of a
CC tissue or cell sample exposed to the compound, and comparing the gene
CC expression profile to a database comprising at least part of the data or
CC information given in the specification. The methods are useful for
CC predicting at least one toxic effect of a compound, predicting the renal
CC progression of a toxic effect of a compound, identifying toxicity markers in tissues or
CC cells exposed to known renal toxin. The genes are useful as toxicity or
CC markers in drug screening and toxicity assays, in monitoring disease or
CC physiological states, or disease progression. This polynucleotide
CC represents a rat DNA sequence relating to the toxic effect database
CC described in the specification. NOTE: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the World Intellectual Property
CC Organization
XX
SQ Sequence 5028 BP; 1289 A; 1273 C; 1058 G; 1408 T; 0 U; 0 Other;

Query Match 20.1%; Score 301.8; DB 10; Length 5028;
Best Local Similarity 66.2%; Pred. No. 2.5e-84;
Matches 451; Conservative 0; Mismatches 227; Indels 3; Gaps 1;
QY 101 TTTCTTGGCAACTGTGTGACAGGCGCTTTAACAGTGTGAGAAATTAAGTTTCATCTCC 160
Db |||||
557 TTTCTGCTCAAAAATGCGCAAGTCTCTCTCCCTGGAGAGTTCCCATCCACAT 616
QY 161 TACTCTCACACAGAGAGAGCGCTTACAAGTGCATACAACAGAGACTGCACCAAGCGCTTT 220
Db |||||
617 TATTTCCACACACAGGAGCGCCCTTCAAGTGTCTTCCAGAGCTGAGTGTGGCAAGCGCTTC 676
QY 221 GTTCTTAAGTACAAATTAACAAGGCACATGGCTACTCTTCTCTGAGAAACCCACCAAG 280
Db |||||
677 GTCTCCAAGTATAGCTGATGAGACACATGGCTACGCACTCTCTCCCAAGAGCGCACCAAG 736
QY 281 TGTAAATTTGAGAAAATGTTTTCACCGGAAGATCATCTGAAGATCACTCCATCAVACA 340
Db |||||
737 TGCACCTATTGTGAAAGACTTTCAACCGGAAGGATCACTGAAGATCACTCCAGACC 796
QY 341 CACGACCTTAACAAGAGACGCTTTAAAGTGCAGAAATGTGGCAAGAACTACAATACCAAG 400
Db |||||
797 CACGATCCCAACAAGATGATCTACGCTTGGAGAGATTTGGCAAGAAATACCAACCATG 856
QY 401 CTGTGATTTAAACGTCACCTGGCCCTTGCATGCCGCAACAAGTGTGACCTCACTGTAAAG 460
Db |||||
857 CTGGCTACAGAGGACATGCTGATGCCCTGCAATTCGGCCAGCAGCGCGATCTCACTGCGGC 916
QY 461 GTATGTTGAAACTTTTGAAGACGAGGAGTGCTTCTGGAGCACTTAAATCTCATGC- 519
Db |||||
917 GTCTGACCCCTGGAGCTGGGAGCACCGAGGTCCTGCTGGACCACTCAAGTCTCACGGC 976
QY 520 --AGCAAGTCTCTGTGGGTAAAGAAAAAAGCAACCAAGTGGCGAAATTTGTGATGC 577

Db 977 GAAGAAAGGCCACACAGCCGCCAGGAGAGAAACACCAAGTCGACCACTGGAGAGA 1036
 QY 578 CGGTTCTACACCCGAAAGGATGTCCGGAGACACATGGTGGTGCACTGGAAGAAAGGAC 637
 Db 1037 TGCTTCTACACCCGAAAGGATGTCCGGAGACACATGGTGGTGCACTGGAAGAAAGGAC 1096
 QY 638 TTCTCTGTCTAGTATTGTGCACAGAGATTTGGGGGAAAGGATCACTGACATGATG 697
 Db 1097 TTCTGTGTCTAGTATTGTGCACAGAGATTTGGGGGAAAGGATCACTGACATGATG 1156
 QY 698 AAGAAGAGTCACAAATCAAGAGCTTCTGAAGGTCAAAACAGACCAAGTGGATTTCCTTGAC 757
 Db 1157 AAGAAGACCACTCCACAGAGCTGTGCAAGAGAGCTTGAAGAGAGAGCTTGAAGAGAGATACCGGGC 1216
 QY 758 CCATTACCTGCAATGTGCT 778
 Db 1217 GGTACCAACCCATTGGCGCT 1237

RESULT 10

AAV18480
 ID AAV18480 standard; cDNA to mRNA; 2790 BP.

AC AAV18480;

DT 18-AUG-1998 (first entry)

DE BOP1 cDNA.

KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;
 KW inducible; alzheimer's disease; nuclear transcription factor; apoptosis;
 KW cell cycle; neuronal disorder; ss.

OS Mus sp.

FH Key Location/Qualifiers

FT CDS 542..2545

FT /*tag= a

FT /product= "BOP1 protein"

XX WO9813489-A1.

XX 02-APR-1998.

XX 22-SEP-1997; 97WO-EP005198.

XX 23-SEP-1996; 96US-00718661.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX (CNRS) CENT NAT RECH SCI.

XX Spengler D, Journot L;

XX WPI; 1998-230701/20.

XX P-PSDB; AAW48760.

PT New isolated tumour suppressor gene - useful for developing products for
 PT use in diagnosis and treatment of tumour(s) or neuronal disorder(s).
 PS Claim 1; Page 72-76; 118pp; English.

CC The present sequence represents the BOP1 cDNA isolated from the mouse
 CC corticotroph pituitary tumour cell line ACT-20 cDNA library. The protein
 CC encoded by the BOP1 cDNA displays a tumour suppressing activity when it
 CC was constitutively and inducibly expressed in tumour cells. The BOP1 cDNA
 CC and the protein it encodes are claimed to be useful in the preparation of
 CC therapeutic compositions, useful for treating, preventing or delaying the
 CC recurrence of a tumour or neuronal disorders, e.g. genetic diseases or
 CC acquired degenerative encephalopathies such as Alzheimer's disease. The
 CC BOP1 protein is also claimed to be able to induce apoptosis resulting in
 CC inhibition of tumour cell growth, to suppress tumour formation, to induce
 CC G1 arrest of the cell cycle and to act as nuclear transcription factor

XX SQ Sequence 2790 BP; 567 A; 783 C; 714 G; 626 T; 0 U; 0 Other;
 Query Match 19.6%; Score 294.4; DB 2; Length 2790;
 Best Local Similarity 64.6%; Pred. No. 4.1e-82;
 Matches 455; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
 QY 78 GTGAACCAACCAAGAAACCTTTCCTTGGCCAACTGTGTGCAAGCCCTTTAACAGTG 137
 Db 528 GAGAAACCAAGGCCATGGCTCCATTCCGTGTCAAAATGTGCAAGTCTTCTGTACCC 587
 QY 138 TTGAGAAATTAAGGTTCACTCTCTACTCTACACAGAGAGAGCCCTTCAAGTGCATAC 197
 Db 588 TGGAGAGTTCACCATTCACATTTATTCACATTCCTCCAGGGAGCGCCATTCAGTGTGCA 647
 QY 198 AACAGACTGCACCAAGGGCTTTGTTTCTAAGTACAAATTTACAAAGGCACATGGCTACTC 257
 Db 648 AGGCTGAGTGTGGCAAGCCTTCGTCTCCAGTATTAAGTGTGAGACATGGCCACAC 707
 QY 258 ATTCTCTGAGAAACCCCAAGTGTAATTTGTGAGAAATGTTTCAACCGAAAGATC 317
 Db 708 ACTCGCCACAGAGATTTCACAGTGTACTCACTGTGAGAAGACATTCACACCGAAGACC 767
 QY 318 ATCTGAAGAAATCACCTCCATACACAGCCCTTAACAAAGAGAGCTTTAAGTGCAGAAAT 377
 Db 768 ACTGAAGAACCACTCCAGACCCAGATCCCAACAAGATCTCTACGGGTGTGACGAT 827
 QY 378 GTGGCAAGAACTACAATACCAAGCTTTGAATTTAAAGCTCACTTGGCTTGCATGCCGAA 437
 Db 828 GCGGCAAGAAAGTACACACCATGCTGGCTACAGAGGACCTGCGCTTGCATCGGCA 887
 QY 438 CAAGTGGTGACCTCACCTGTAAAGTATGTTTGCAAACTTTTGAAGACGCGGAGTGCTTC 497
 Db 888 GCAATGGCGATCTCACCTGTGGGTGTGCACCTGGAGCTGGGAGACCCGAGGTCCTGC 947
 QY 498 TGGAGCACTTAAATCTCATGC--AGGCAAGTCTCTGTGGGGTTAAAGAAAAAAGC 554
 Db 948 TGGACCACCTCAAGTCTCACGGGAAGAAAGGCCAACCCAGCCACCCAGGAGAAGAAAT 1007
 QY 555 ACCAGTGCAGCAATTGTGATCGCGGTTCTACACCCGAAAGATGTCTCGGAGACACATGG 614
 Db 1008 ACCAGTGCACCACTGTGATAGATGCTTCTACACCCGAAAGATGTGCTGCCACCTGG 1067
 QY 615 TGGTGCACACTGGAAGAAAGACTTCTCTCTAGTATTTGTGCACAGAGATTTGGGCGAA 674
 Db 1068 TGGTCCACACAGGATGCAAGGACTTCTCTGTCTAGTCTGTGCCAGAGATTTGGGCGCA 1127
 QY 675 AGGATCACCTGACTCGACATATGAAGAGAGTCAATCAAGACTTCTGAAGTCAAAA 734
 Db 1128 AAGACCACCTCACTCGTCCACCAAGAGAGACCCCACTCCAGGAGCTGATGCAGAGAATA 1187
 QY 735 CAGAACCAAGTGGATTTCCTTGGACCCCAATTTACCTGCAATGTGCT 778
 Db 1188 TGCAGGAGGAGATTACACAGAGCAATTTCCAACTCATTTGGCGCT 1231

RESULT 11

ABN86526

ID ABN86526 standard; cDNA; 3975 BP.

XX AC ABN86526;

XX AC ABN86526;

XX 21-OCT-2002 (first entry)

DE Nucleotide sequence of mouse ZAC zinger finger protein cDNA.

XX Cardiovascular; Flt-1; CD44; Lot-1; AA892598; Mrp-1; cardiant; mouse;

XX cerebroprotective; antiarteriosclerotic; ZAC; zinger finger protein; ss.

OS Mus musculus.

XX WO200238794-A2.

PN

XX

615 TGGTGACACCTGGAGAAAGGACTTCTCTCTGAGTATTGTGCACAGAGATTGGCGCAA 674
775 TGGTCCACACAGGATGCAAGGACTTCTCTCTGAGTATTGTGCACAGAGATTGGCGCAA 834
675 AGGATCACTGACCTGACATATGAAGAAGAGTCAATCAAGAGCTTCTGAAGTCAAAA 734
835 AAGACCACTCTCTGTCACACCAAGAGAGCCACTCCAGGAGTGTGCAAGAGAATA 894
735 CAGAACCACTGAGTCTCTGACCCATTACCTGCAATGTGTCT 778
895 TGCAGGAGGAGATTACCAAGAGCAATTTCCAACTCATTGCGCCT 938

RESULT 12
ADI26094
ID ADI26094 standard; cDNA; 3975 BP.
XX AC ADI26094;
XX DT 22-APR-2004 (first entry)
XX DE Human cDNA encoding protein that promotes STAT6 activation #30.
XX KW ss; gene; human; signal transducer and activator of transcription 6;
XX KW STAT6; immunogen; STAT6 activation; allergy; inflammation;
XX KW autoimmune disease; diabetes; hyperlipidaemia; infection; cancer;
XX KW Th1 hyperactive disease; rheumatoid arthritis; osteoarthritis;
XX KW systemic lupus erythematosus; sepsis; asthma; allergic rhinitis;
XX KW ischaemic heart disease; subarachnoid haemorrhage; viral hepatitis; AIDS.
XX OS Homo sapiens.
XX PN WC2003104277-A2.
XX PD 18-DEC-2003.
XX PF 05-JUN-2003; 2003WO-JP007123.
XX PR 05-JUN-2002; 2002JP-00164257.
XX PR 06-JUN-2002; 2002US-0385912P.
XX PR 26-DEC-2002; 2002JP-00377326.
XX PR 27-DEC-2002; 2002US-0436467P.
XX PR 15-MAY-2003; 2003JP-00137505.
XX PR 16-MAY-2003; 2003JP-0470836P.
XX PA (ASAH) ASahi KASEI KK.
XX PI Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;
XX WPI; 2004-122214/12.
XX P-PSDB; ADI26095.
XX PT New signal transducer and activator of transcription 6 activation
XX PT promoting purified protein, for diagnosing and treating disease
XX PT associated with activation/inhibition of transcription factor e.g.
XX PT diabetes and cancer.
XX PS Claim 4; SEQ ID NO 59; 1368pp; English.
XX CC The invention relates to a purified protein promoting signal transducer
XX CC and activator of transcription 6 activation (STAT6). The protein is
XX CC useful for producing an antibody, which involves administering the
XX CC protein or its epitope-bearing fragments to a non-human animal as an
XX CC antigen. The nucleic acid is useful for diagnosing a disease or
XX CC susceptibility to a disease related to expression or activity of the
XX CC protein. A transformant expressing the protein is useful for screening
XX CC compounds which inhibit or promote STAT6 activation. A transformant
XX CC expressing the protein is useful for producing a pharmaceutical
XX CC composition. Compositions, antibodies and antisense molecules are useful
XX CC for the treating a disease associated with STAT6 activation such as
XX CC allergic diseases, inflammation, autoimmune diseases, diabetes,
XX CC hyperlipidaemia, infections disease and cancers. Compositions are useful
XX CC for treating disease associated with STAT6 activation and/or prevention

16-MAY-2002.
08-NOV-2001; 2001WO-US046816.
09-NOV-2000; 2000US-0247457P.
(BGHM) BRIGHAM & WOMENS HOSPITAL INC.
Lee RT;
WPI; 2002-590446/63.
Diagnosing cardiovascular conditions e.g. stroke, myocardial infarction
in a subject, by detecting expression of a nucleic acid molecule such as
Fit-1, CD44, Lot-1, AA892598 and Mrg-1 in biological sample from subject.
Disclosure; Page 108-109; 113pp; English.
The invention relates to diagnosing a cardiovascular condition
characterised by aberrant expression of a nucleic acid molecule (I) such
as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 or its expression product. One
method involves contacting a biological sample from a subject with an
agent which specifically binds to (I), its expression product or a
fragment of an expression product and measuring the amount of bound
agent. The method is useful for diagnosing a cardiovascular condition
such as myocardial infarction, stroke, arteriosclerosis, heart failure,
and cardiac hypertrophy. Methods useful for determining the stage of
cardiovascular condition and pharmaceutical compositions for treating the
above cardiovascular conditions are also provided. The present sequence
represents a mouse ZAC zinger finger protein cDNA
Sequence 3975 BP; 1010 C; 880 G; 1084 T; 0 U; 0 Other;
Query Match 19.6%; Score 294.4; DB 6; Length 3975;
Best Local Similarity 64.6%; Pred. No. 5e-82;
Matches 455; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
78 GTGAACCAACCAAGAAAATCTTCTTGGCAACTGTGTGCAAGGCCCTTTAACAAGTG 137
235 GAGAACCAAGGCCATGGCTCCATTCCGCTGTCAAAAATGGCAAGTCTTCTGTACCC 294
138 TTGAGAAATTAAGGTTCACTCTCTACACAGGAGAGAGGCCCTTACAGTGCATAC 197
295 TGGAGAAGTTACCATTCACAAATATTTCCCACTCCAGGAGCGCCCATTCAGTGCTGA 354
198 AACAGAAGTGCACCAAGGCCCTTTGTTCTTAAGTACAAATTAACAAGGCACATGGCTACTC 257
355 AGGCTGAGTGTGGCAAGCCCTTCTGTCCTCAAGTATAGCTGATGAGACACATGGCCACAC 414
258 ATTCTCTGGAAGAACCAAGTGTATTTCTGAGAAATATTTTCAACCGGAAGATC 317
415 ACTGCCACAGAAGATTCACCAAGTGTATCTCACTGTGGAAGACATCAACCGGAAGACC 474
318 ATCTGAAGATCACTCCATACACAGCCCTTAAACAAGAGACGTTAAGTTCGGAAGAT 377
475 ACCTGAAGAACCACTCCAGACCCACGATCCCAAGATCTCTACGGCTGTGACGATT 534
378 GTGGCAAGAACTACAAATCAAGCTTGGATTAAACGTCATTGGCTTGCATGCGGCA 437
535 GCGGCAAGAGTACACACCATGCTGGGTACAAAGAGGACCTGGCCCTGCACTCGGGA 594
438 CAAAGTGTGACCTCACTGTAAGTATGTTTGCAAACTTTGAAGACGCGAGTGCTTC 497
595 GCAATGGCGATCTCACTGTGGGGTGTGACCCCTGGAGCTGGGAGCACCGAGTCTGC 654
498 TGGAGACCTTAAATCTCATGCG --- AGGCAAGTCTGCTGGTGGGTAAAGAAAAGAC 554
655 TGAACCACTCAGTCTCAGCGGAGAGAAAGGCCAACAGGACACCGAGGAGAGAAAT 714
555 ACCAGTGGCAACATTTGATTCGCGGTTTCTACCCGAAAGGATGTCGGAGACATGG 614
715 ACCAGTGGCAACATTTGATTCGCGGTTTCTACCCGAAAGGATGTCGGTGGCCACTGG 774

Db 411 AGCACGAGGTGCTACTGGACCACTCAAGCCCATCGGAGAGAGAGCCCTTAGCGGA 470
 Qy 539 GTTAAAGAAAAAGCACCAGTGGGAAATTTGTGATCGCCGGTCTTACACCGGAAAGAT 598
 Db 471 ACCAAGGAAAAAGCACCAGTGGGAAATTTGTGATCGCCGGTCTTACACCGGAAAGAT 530
 Qy 599 GTCCGGAGACATGGTGGTGCACACTGGAAGAGAGACTTCTCTGTCAGTATTGTGCA 658
 Db 531 GTGCACGCCACCTGGTGGTCCACAGAGATCAAGGACTTCTGTGCCAGTTCTGTGCC 590
 Qy 659 CAGAGATTGGCGGAAAGATCACCTGACTGCACATATGAAGAGAGTCAACATCAAGAG 718
 Db 591 CAGAGATTGGCGGAAAGATCACCTGACTGCACATATGAAGAGAGTCAACATCAAGAG 650
 Qy 719 CTCTGAAAGGTCAAAACAGAACAGTGGGATTTCCTTGACCCATT 762
 Db 651 CTGATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGCACCTT 694

RESULT 14
 ID ABT11028 standard; cdna; 2561 BP.
 XX ABT11028;
 AC ABT11028;
 XX 04-DEC-2002 (first entry)
 XX Human breast cancer associated coding sequence SEQ ID NO: 1162.
 XX Human; breast specific gene; breast cancer; differential expression;
 KW cytosstatic; gene therapy; gene; ss.
 XX Homo sapiens.
 XX WO200259271-A2.
 XX 01-AUG-2002.
 XX 25-JAN-2002; 2002WO-US002176.
 XX 25-JAN-2001; 2001US-0263757P.
 XX 23-APR-2001; 2001US-0286090P.
 XX 23-MAY-2001; 2001US-0292517P.
 XX (GENE-) GENE LOGIC INC.
 XX Orr MS, Nation M, Diggans JC, Zeng W;
 XX WPI; 2002-674803/72.

Diagnosing breast cancer in a patient comprises detecting the level of gene expression in cell or tissue samples, where a differential gene expression is indicative of breast cancer.
 Claim 1; SEQ ID NO 1162; 260pp + Sequence Listing; English.
 The present invention relates to methods of diagnosing breast cancer in a patient, which comprise detecting the level of expression in a tissue sample of two or more genes selected from those shown in ABO9867-ABT1112, where a differential expression of the genes indicates breast cancer. The methods are useful in diagnosing, treating, detecting the progression, and in monitoring treatment of breast cancer in patients. The methods are also useful as a screening tool for agents that modulate the onset or progression of breast cancer. The breast cancer genes may be used as diagnostic markers for the prediction or identification of the malignant state of breast tissue, for confirming the type and progression of cancer, and for drug screening and assays. The present sequence is a coding sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub.published_pct_sequences
 Sequence 2561 BP; 704 A; 624 C; 516 G; 717 T; 0 U; 0 Other;

Query Match 18.5%; Score 277.6; DB 6; Length 2561;
 Best Local Similarity 71.9%; Pred. No. 8.7e-77;
 Matches 377; Conservative 0; Mismatches 144; Indels 3; Gaps 1;
 Qy 242 AGGCACATGGCTACTCATTCTCTCTGAGAAAAACCCCAAGTGTAAATTATTTGAGAAATG 301
 Db 171 AGGCATATGGCTATCCCATTTCTCCAGAAATCTCCACAGTGTCTCACTGTGAGAAAGC 230
 Qy 302 TTTCACCGGAAAGATCATCTGAGAAATCACCTCCATACACAGCCCTTAAACAGAGAGC 361
 Db 231 TTCAACCGGAAAGACCACTGAAATACCACTCCAGACCCAGACCCCAAAATGGCC 290
 Qy 362 TTTAAGTGGGAAAGATGTGGCAAGTACCAATACCAAGCTTGGATTAAACGTCACTTG 421
 Db 291 TTTGGGTGTGAGAGTGTGGGAAAGATCAACACCATCTGGCTATTAAGAGGACCTG 350
 Qy 422 GCCTTGCATGCCGCAACAAAGTGGTGACCTCACCTGTAAAGTATGTTTGAAACTTTGAA 481
 Db 351 GCCCTCCATGCCGCCAGCAGTGGGACCTCACCTGTGGGTCTGTGCCCTGAGCTAGGG 410
 Qy 482 AGCACGGGAGTCTTCTGAGACACCTTAAATCTCATGC--AGGCAAGTCTGTGGG 538
 Db 411 AGCACGGGAGTCTTCTGAGACACCTTAAATCTCATGC--AGGCAAGTCTGTGGG 470
 Qy 539 GTTAAAGAAAAAGCACCAGTGGGAAAGATGTGGGAAAGATGTGGGAAAGATGTGGG 598
 Db 471 ACCAAGGAAAAAGCACCAGTGGGAAAGATGTGGGAAAGATGTGGGAAAGATGTGGG 530
 Qy 599 GTCCGGAGACATGGTGGTGCACACTGGAAGAAAGAGACTTCTCTGTCAGTATTGTGCA 658
 Db 531 GTCCGAGCCACCTGGTGGTCCACACAGGATCAAGGACTTCTGTGCCAGTTCTGTGCC 590
 Qy 659 CAGAGATTGGCGGAAAGATCACTGACATATGAAGAGAGTCAACATCAAGAG 718
 Db 591 CAGAGATTGGCGGAAAGATCACTGACATATGAAGAGAGTCAACATCAAGAG 650
 Qy 719 CTCTGAGGTCAAAACAGAGACCTGAGTGTTCCTTGCACCCATT 762
 Db 651 CTGATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGCACCTT 694

RESULT 15
 ID AAZ33549 standard; cdna; 2738 BP.
 XX AAZ33549;
 AC AAZ33549;
 XX 08-DEC-1999 (first entry)
 XX Human breast tumour-associated EST 9.
 DE Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 KW medicaments; gene therapy; treatment; fat metabolism; ss.
 XX Homo sapiens.
 OS
 XX DE19813835-A1.
 XX 23-SEP-1999.
 XX 20-MAR-1998; 98DE-01013835.
 XX 20-MAR-1998; 98DE-01013835.
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
 XX WPI; 1999-528979/45.
 XX P-PSDB; AAY48468.
 XX Human nucleic acid sequences and protein products from normal breast

Search completed: November 8, 2004, 19:16:59
Job time : 790 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2004, 10:22:19 ; Search time 5072 Seconds
(without alignments)
10783.919 Million cell updates/sec

Title: US-09-242-772-116_COPY_480_1980

Perfect score: 1501

Sequence: 1 gatggccactgtcattctcg.....gtttccatcaagcttttccag 1501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1498.4	99.8	1503	9 AY420150	AY420150 Homo sapi
2	1455.4	97.0	1503	9 AY420151	AY420151 Pan trogl
3	1260	83.9	1500	9 AY420152	AY420152 Mus muscu
4	805.8	53.7	1185	5 BX424854	BX424854 BX424854
5	793.6	52.9	864	5 BX370812	BX370812 BX370812
6	662.2	44.1	736	1 AL036879	AL036879 DRFZP564P
7	543.4	36.2	734	6 CA411972	CA411972 UI-H-EZO-
8	434	28.9	492	8 B94684	B94684 CIT-HSP-217
9	421.6	28.1	642	5 BU210438	BU210438 604156514
10	396.2	26.4	683	6 BY179727	BY179727 BX719727
11	388.2	25.9	4767	3 AK040626	AK040626 Mus muscu
12	373.8	24.9	1047	5 BX402039	BX402039 BX402039
13	352.2	23.5	365	8 AQ125617	AQ125617 HS 2170 A
14	347	23.1	804	5 BU260281	BU260281 603501940
15	346.2	23.1	386	8 B99080	B99080 CIT-HSP-228
16	333.8	22.2	807	5 BU186849	BU186849 AGENCOURT
17	332.4	22.1	844	7 CN533596	CN533596 UI-M-HSO-
18	325.8	21.7	845	5 BU172561	BU172561 AGENCOURT
19	312.6	20.8	743	7 CN537373	CN537373 UI-M-HSO-
20	311	20.7	786	7 CK637889	CK637889 UI-M-HSO-
21	308.6	20.6	553	7 CN534810	CN534810 UI-M-HSO-
22	304.8	20.3	587	7 CN538026	CN538026 UI-M-HSO-
23	300.4	20.0	797	9 CN50311Y	AL245023 Tetraodon
24	300.2	20.0	546	7 CN312865	CN312865 170004177

25	299.6	20.0	822	4 BI820689	BI820689 603034475
26	293.8	19.6	901	5 BU161452	BU161452 AGENCOURT
27	290.4	19.3	999	9 CNS03809	AL232866 Tetraodon
28	288.8	19.2	667	5 BX642593	BX642593 DRFZP5686N
29	281	18.7	1098	1 AL523896	AL523896 AL523896
30	279.2	18.6	913	1 AL542380	AL542380 AL542380
31	278.2	18.5	854	1 AU136661	AU136661 AU136661
32	273.2	18.2	720	7 CN312867	CN312867 170004245
33	271.8	18.1	617	7 CN665887	CN665887 A0832806-
34	266.8	17.8	563	6 CA750971	CA750971 UI-M-ROO-
35	266	17.7	838	7 CK601489	CK601489 AGENCOURT
36	262	17.5	711	7 CN392814	CN392814 170005315
37	261.8	17.4	600	4 BI989866	BI989866 4052-32 M
38	256.4	17.1	582	2 AW952782	AW952782 EST364852
39	250.2	16.7	700	5 BP761624	BP761624 BP761624
40	249.2	16.6	668	7 CF360369	CF360369 821696 MA
41	248.4	16.5	669	6 CB017609	CB017609 pgnic.pk0
42	246.2	16.4	1075	5 BX356755	BX356755 BX356755
43	240.8	16.0	855	4 BI826943	BI826943 603075831
44	238.4	15.9	582	6 CA949670	CA949670 IQ23F04.Y
45	238.2	15.9	964	1 AL526858	AL526858 AL526858

ALIGNMENTS

RESULT 1
AY420150
LOCUS
DEFINITION
Homo sapiens PLAG1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION
AY420150
VERSION
AY420150.1 GI:39776107
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
AUTHORS
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302
REFERENCE
2 (bases 1 to 1503)
AUTHORS
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Direct Submission
JOURNAL
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
Location/Qualifiers
source
1..1503
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1503
/gene="PLAG1"
/locus_tag="HCW7128"

Query Match 99.8%; Score 1498.4; DB 9; Length 1503;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1499; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ATGGCCACTGTCATTCCTGGTGAATTCAGAGTAAGAGATACCCAGAAAGTCCCTTCA 61
|||||

Db 1 ATGGCCACTGTCATTCCTGGTGATTTTGTCAAGTAAGAGATACCAGAAAGTCCCTTCA 60
Qy 62 GGGAAACGTTAAGCGTGGTGAACCAACCAAGAAAAAATTTCTTCCCAACTGTGTAC 121
Db 61 GGGAAACGTTAAGCGTGGTGAACCAACCAAGAAAAAATTTCTTCCCAACTGTGTAC 120
Qy 122 AAGGCTTTAAGCTGTGTGAGAAATTAAGGTTCACTCTACTCTCAACAGGAGAGAGG 181
Db 121 AAGGCTTTAAGCTGTGTGAGAAATTAAGGTTCACTCTACTCTCAACAGGAGAGAGG 180
Qy 182 CCCTACAGTGCATACACAGACTGCACCAAGGCTTTGTTTCTTAAGTCAAAATTAACA 241
Db 181 CCCTACAGTGCATACACAGACTGCACCAAGGCTTTGTTTCTTAAGTCAAAATTAACA 240
Qy 242 AGGCATAGGCTACTCATTTCTCTGAGAAACCCCAAGTGTAATATTTGAGAAATG 301
Db 241 AGGCATAGGCTACTCATTTCTCTGAGAAACCCCAAGTGTAATATTTGAGAAATG 300
Qy 302 TTTCAACGGAAGATCACTGAAGATCACTCCATACACAGCCCTTAACAAAGAGAGG 361
Db 301 TTTCAACGGAAGATCACTGAAGATCACTCCATACACAGCCCTTAACAAAGAGAGG 360
Qy 362 TTTAAGTGGAGAGATGTGGCAAGACTACAATACCAAGCTTGGATTAAAGCTCACTTG 421
Db 361 TTTAAGTGGAGAGATGTGGCAAGACTACAATACCAAGCTTGGATTAAAGCTCACTTG 420
Qy 422 GCCTTCATGCCGCAACAGTGGTGACCTCACTGTAAGTATGTTTGCACCTTTTGA 481
Db 421 GCCTTCATGCCGCAACAGTGGTGACCTCACTGTAAGTATGTTTGCACCTTTTGA 480
Qy 482 AGCAGGGAGTCTTCTGGAGCACTTAAATCTCATGCGGCAAGTCTGTGGGGGT 541
Db 481 AGCAGGGAGTCTTCTGGAGCACTTAAATCTCATGCGGCAAGTCTGTGGGGGT 540
Qy 542 AAGAAAAAAGCACCAGTCCGCAATTTGTGATCGCGGTCTTACACCCGAAAGGATGTC 601
Db 541 AAGAAAAAAGCACCAGTCCGCAATTTGTGATCGCGGTCTTACACCCGAAAGGATGTC 600
Qy 602 CGGAGACACATGTTGGTGACACTGGAAGAAAGACTTCTCTGTCACTAGTATGTCACAG 661
Db 601 CGGAGACACATGTTGGTGACACTGGAAGAAAGACTTCTCTGTCACTAGTATGTCACAG 660
Qy 662 AGATTGGCGAAGAGTACCTGACCTCGACATATGAAGAGAGTCAATCAAGAGCTT 721
Db 661 AGATTGGCGAAGAGTACCTGACCTCGACATATGAAGAGAGTCAATCAAGAGCTT 720
Qy 722 CTGAAGTCAAAACAGACAGTGGATTTCTTCAACCATTTACCTGCAATGTGTCTGTG 781
Db 721 CTGAAGTCAAAACAGACAGTGGATTTCTTCAACCATTTACCTGCAATGTGTCTGTG 780
Qy 782 CCTATAAAGACGAGCTCCTTCCGGTGATGTCCTTACCTTCCAGTGAAGTCAATCAAG 841
Db 781 CCTATAAAGACGAGCTCCTTCCGGTGATGTCCTTACCTTCCAGTGAAGTCAATCAAG 840
Qy 842 CCATTCAACACCTTTCAGTAAACCTCTACACACTCCATTTCACTGTCATGAGAGC 901
Db 841 CCATTCAACACCTTTCAGTAAACCTCTACACACTCCATTTCACTGTCATGAGAGC 900
Qy 902 TCGGATCTGCCACCAAAATGATCAACTTTTACCTTTGGGAATGACATGCCAATAGAT 961
Db 901 TCGGATCTGCCACCAAAATGATCAACTTTTACCTTTGGGAATGACATGCCAATAGAT 960
Qy 962 AFGGACATGTTTCATCCCTCTCACACCTTTCTTCAATATCCGTTCACTTCACTCA 1021
Db 961 AFGGACATGTTTCATCCCTCTCACACCTTTCTTCAATATCCGTTCACTTCACTCA 1020
Qy 1022 TATGCAATTTCTATTTCTGAAAAAGAACAGCCATTAAAGGGGAAATTCAGAGTTTACCTG 1081
Db 1021 TATGCAATTTCTATTTCTGAAAAAGAACAGCCATTAAAGGGGAAATTCAGAGTTTACCTG 1080
Qy 1082 ATGGAGTTAAGAGTGGCGTCCCTCTTATCCCAAGATTTCTCAAGCATGCTCATCT 1141
Db 1081 ATGGAGTTAAGAGTGGCGTCCCTCTTATCCCAAGATTTCTCAAGCATGCTCATCT 1140

Qy 1142 AAGCTAGGTTGGATCTCAGATTGGTCCCTAGATGATGGTGCAGAGAGCTTCCCTA 1201
Db 1141 AAGCTAGGTTGGATCTCAGATTGGTCCCTAGATGATGGTGCAGAGAGCTTCCCTA 1200
Qy 1202 TCCAAAAGCTCTATCTCCATCAGTGACCCCTAAACACACAGATTGGATTTTCTCAG 1261
Db 1201 TCCAAAAGCTCTATCTCCATCAGTGACCCCTAAACACACAGATTGGATTTTCTCAG 1260
Qy 1262 TTGTTTAAATTTCAATCCTTTAAATGGTCTCTCTATATCTCTATCAGTGGGAGCCTT 1321
Db 1261 TTGTTTAAATTTCAATCCTTTAAATGGTCTCTCTATATCTCTATCAGTGGGAGCCTT 1320
Qy 1322 GGAATGAGCTATTCACAGAGAGAGACATTTCTTGTTCCTTCCAGCTCCCAACAAACA 1381
Db 1321 GGAATGAGCTATTCACAGAGAGAGACATTTCTTGTTCCTTCCAGCTCCCAACAAACA 1380
Qy 1382 CAGGATCTTCAGGATCTCTGCAAACTATAGGCTTGGTCTCTGCACTCAGTCTCAGCA 1441
Db 1381 CAGGATCTTCAGGATCTCTGCAAACTATAGGCTTGGTCTCTGCACTCAGTCTCAGCA 1440
Qy 1442 GCTTTCCACAGAGTTTAAAGCAAGTACACCTCCAGCTTCCATCAAGCTTTTCTCAG 1501
Db 1441 GCTTTCCACAGAGTTTAAAGCAAGTACACCTCCAGCTTCCATCAAGCTTTTCTCAG 1500

RESULT 2
AY420151 1503 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes PLAG1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY420151 GI:39776108
VERSION AY420151.1
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 1503)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
Gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1503)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source
1..1503
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1503
/gene="PLAG1"
/locus_tag="HCM7128"
gene
ORIGIN
Query Match 97.0%; Score 1455.4; DB 9; Length 1503;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1456; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Qy 2 ATGGCCACTGTCATTCCTGGTGATTTTGTCAAGTAAGAGATACCAGAAAGTCCCTTCA 61

Db 1 ATGGCCACTGTCAUCCCTGGTGATTTGTCAAGATAGAGATACCCAGAAAGTCCCTTCA 60
 QY 62 GGGAAACGTAGCGTGTGAACCAACCAAGAAACCAATTTCTTGGCAACTGTGTGAC 121
 Db 61 GGGAAACGTAGCGTGTGAACCAACCAAGAAACCAATTTCTTGGCAACTGTGTGAC 120
 QY 122 AAGSCCTTTAAACAGTGTGTGAGAAATTAAGAGTTCACCTCTACTCTCACAGGAGAGG 181
 Db 121 AAGSCCTTTAAACAGTGTGTGAGAAATTAAGAGTTCACCTCTACTCTCACAGGAGAGG 180
 QY 182 CCCTAAGTGTATACAAACAGACTGCACCAAGGCTTTTGTCTTAAGTCAAAATACAA 241
 Db 181 CCCTAAGTGTATACAAACAGACTGCACCAAGGCTTTTGTCTTAAGTCAAAATACAA 240
 QY 242 AGSCACATGCTACTCTTCTCTGAGAAACCCCAAGTGTAAATTTATGTGAGAAATG 301
 Db 241 AGSCACATGCTACTCTTCTCTGAGAAACCCCAAGTGTAAATTTATGTGAGAAATG 300
 QY 302 TTTACCGGAAAGATCATCTGAAGATCACTCCATACACACACCCCTTAACAAAGAGCG 361
 Db 301 TTTACCGGAAAGATCATCTGAAGATCACTCCATACACACACCCCTTAACAAAGAGCG 360
 QY 362 TTTAAGTGTGAAGATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAACGTCACTTG 421
 Db 361 TTTAAGTGTGAAGATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAACGTCACTTG 420
 QY 422 GCCTTGATGTCGCGCAACAGTGTGACCTCACTGTAAAGTATGTTTGCAAACTTTTGAA 481
 Db 421 GCCTTGATGTCGCGCAACAGTGTGACCTCACTGTAAAGTATGTTTGCAAACTTTTGAA 480
 QY 482 AGCAGCGGAGTCTTCTTGAGCAGCACTTAAATCTCATGAGCAAGTCTGTCTGGTGGGTT 541
 Db 481 AGCAGCGGAGTCTTCTTGAGCAGCACTTAAATCTCATGAGCAAGTCTGTCTGGTGGGTT 540
 QY 542 AAAGAAAAAAGACCAAGTGTGCAATGTGATCGCCGGTCTTACACCGAAGATGTC 601
 Db 541 AAAGAAAAAAGACCAAGTGTGCAATGTGATCGCCGGTCTTACACCGAAGATGTC 600
 QY 602 CGGAGACATGTTGGTGTGACACTGTGAAGAAAGGACTTCTCTGTGATTTGTGCACAG 661
 Db 601 CGGAGACATGTTGGTGTGACACTGTGAAGAAAGGACTTCTCTGTGATTTGTGCACAG 660
 QY 662 AGATTTGGGCAAGAGTCACTGACTCGCATATGAAGAGAGTCAATCAAGAGTT 721
 Db 661 AGATTTGGGCAAGAGTCACTGACTCGCATATGAAGAGAGTCAATCAAGAGTT 720
 QY 722 CTGAAGTCAAAAAGACCAAGTGTGATTTCTTGCACCAATTTACCTGCAATGTCTGTG 781
 Db 721 CTGAAGTCAAAAAGACCAAGTGTGATTTCTTGCACCAATTTACCTGCAATGTCTGTG 780
 QY 782 CCTATAAAGACGAGTCTCTTCCGGTGTGATTTCTTGCACCAATTTACCTGCAATGTCTGTG 841
 Db 781 CANNATAAAGACGAGTCTCTTCCGGTGTGATTTCTTGCACCAATTTACCTGCAATGTCTGTG 840
 QY 842 CCATTCACAAACTTTGCAAGTGTAAACCTCTACAACTCCATTTAGTCAAGTGTATCAAG 901
 Db 841 CCATTCACAAACTTTGCAAGTGTAAACCTCTACAACTCCATTTAGTCAAGTGTATCAAG 900
 QY 902 TCGGATCTGCCCAACCAATGATCAACACTTTTACCTTTGGGAATGATGCCCAATAGAT 961
 Db 901 TCGGATCTGCCCAACCAATGATCAACACTTTTACCTTTGGGAATGATGCCCAATAGAT 960
 QY 962 ATGGACACTGTTTCATCTCTGACCACTTTCTTCAAAATATCCGTTCAGTCTCACTCA 1021
 Db 961 ATGGACACTGTTTCATCTCTGACCACTTTCTTCAAAATATCCGTTCAGTCTCACTCA 1020
 QY 1022 TATGCAATTTCTATCTCTGAAAGAAACAGCAATTAAGGGGAAATTTAGAGTTACCTG 1081
 Db 1021 TATGCAATTTCTATCTCTGAAAGAAACAGCAATTAAGGGGAAATTTAGAGTTACCTG 1080
 QY 1082 ATGGAGTTCAAGGTTGGGCTGCTTCTTCAATCCCAAGATTTCTCAAGCATCTCATCT 1141
 Db 1081 ATGGAGTTCAAGGTTGGGCTGCTTCTTCAATCCCAAGATTTCTCAAGCATCTCATCT 1140

QY 1142 AAGCTAGGTTGGATCCTCAGATGGTCCCTAGATGGTGGAGAGACTCTCCCTA 1201
 Db 1141 AAGCTAGGTTGGATCCTCAGATGGTCCCTAGATGGTGGAGAGACTCTCCCTA 1200
 QY 1202 TCCAAAAGCTCTACTCTCCATCAGTACCCCTTAAACACACAGCATTTGGATTTCTCAG 1261
 Db 1201 TCCAAAAGCTCTACTCTCCATCAGTACCCCTTAAACACACAGCATTTGGATTTCTCAG 1260
 QY 1262 TTGTTAAATTTTACATCCTTTAAATGGTCCCTCTATAATCTCTATCAGTGGGAGGCTT 1321
 Db 1261 TTGTTAAATTTTACATCCTTTAAATGGTCCCTCTATAATCTCTATCAGTGGGAGGCTT 1320
 QY 1322 GGAATGAGCTATTTCCAGGAGAACACATTTCTTCTTCTCCAGCTCCCAACAAACA 1381
 Db 1321 GGAATGAGCTATTTCCAGGAGAACACATTTCTTCTTCTCCAGCTCCCAACAAACA 1380
 QY 1382 CAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGACGA 1441
 Db 1381 CAGGATCTTCAGGATCTCTCAACACTATAGGCTTGGGTCTCTGCACTCACTGTGACGA 1440
 QY 1442 GCTTTCACAGCAGTTTAAAGCACAAGTACCACTCCAGCTCCCAAGCTTTTCCAGTTTCAG 1501
 Db 1441 GCTTTCACAGCAGTTTAAAGCACAAGTACCACTCCAGCTCCCAAGCTTTTCCAGTTTCAG 1500

RESULT 3
 AY420152
 LOCUS
 DEFINITION
 Mus musculus PLAG1 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION
 AY420152
 VERSION
 AY420152.1
 GI:39776109
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1500)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)
 JOURNAL
 PUBMED
 14671302
 REFERENCE
 2 (bases 1 to 1500)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Adams, M.D. and Cargill, M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT
 This sequence is made by sequencing genomic exons and ordering them
 based on alignment.
 FEATURES
 source
 1. .1500
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1. .>1500
 /gene="PLAG1"
 /locus_tag="HMC7128"

ORIGIN
 Query Match 83.9%; Score 1260; DB 9; Length 1500;
 Best Local Similarity 90.5%; Pred. No. 0;
 Matches 1357; Conservative 0; Mismatches 140; Indels 3; Gaps 1;
 QY 2 ATGGCCACTGTCAUCCCTGGTGATTTGTCAAGATAGAGATACCCAGAAAGTCCCTTCA 61
 |||||


```

QY 235 ATTACAAAGGCACATGCTACTCTCTCTCTGAGAAAACCCACAAAGTGTATTAATTATTGTA 294
Db 275 AGTAATCAGCAGCATGCTACTCTCTCTCTGAGAAAACCCACAAAGTGTATTAATTATTGTA 334
QY 295 GAAATATGTTTACCGGAAAGATCATCTGGAAGAATCACTCCATACACACCAACCTTAACAA 354
Db 335 GAAATATGTTTACCGGAAAGATCATCTGGAAGAATCACTCCATACACACCAACCTTAACAA 394
QY 355 AGAGAGCTTTTAAGTGGAGAGATGTTGGCAAGAACTACAATACCAAGCTTGGATTTAAACG 414
Db 395 AGAGAGCTTTTAAGTGGAGAGATGTTGGCAAGAACTACAATACCAAGCTTGGATTTAAACG 454
QY 415 TCACCTTGGCTTTCATGCGCGCAACAGTGGTGAACCTCACTGTAAGGATGTTTGAACAC 474
Db 455 TCACCTTGGCTTTCATGCGCGCAACAGTGGTGAACCTCACTGTAAGGATGTTTGAACAC 513
QY 475 TTTTGAAGCAGCGGAGTCTCTGAGGACACCTTAAATCTCATGCGAGCAAGTCTGCTGG 534
Db 514 TTTTGAAGCAGCGGAGT-CTTCTGGAGCACCTTAAATCTCATGCGAGCAAGTCTGCTGG 572
QY 535 TGGGCTTAAAGAAAAGACACAGTGGCAACATTTGTATGTCGCGGTTCTACACCCGAAA 594
Db 573 TGGGCTTAAAGAAAAGACACAGTGGCAACATTTGTATGTCGCGGTTCTACACCCGAAA 632
QY 595 GGATGTCGCGAGACACATGTTGGTGCACACTGGAGAAAGAGACTTCTCTGTGCAATTTG 654
Db 633 GGATGTCGCGAGACACATGTTGGTGCACACTGGAGAAAGAGACTTCTCTGTGCAATTTG 692
QY 655 TGCACAGAGATTTGGCGGAAAGGATCACTGACTCGACATATGAGAGAGAGTCAACAATCA 714
Db 693 TGCACAGAGATTTGGCGGAAAGGATCACTGACTCGACATATGAGAGAGAGTCAACAATCA 752
QY 715 AGAGCTTCTGAAGTCAAAAAGACAGACAGTGGATTTCTTGAACCAATTTACTCTGCAATGT 774
Db 753 AGAGCTTCTGAAGTCAAAAAGACAGACAGTGGATTTCTTGAACCAATTTACTCTGCAATGT 812
QY 775 GTCTGTGCTATAAAGACGAGCTCTCTCCGATGATGCTTACTTCCAGTGAACCTGTT 834
Db 813 GTCTGTGCTATAAAGACGAGCTCTCTCCGATGATGCTTACTTCCAGTGAACCTGTT 872
QY 835 ATCAAAGCCATTACAAACACTTTTGCAGTAAACCTCTACAACTCTTACCACTCCATTTG 894
Db 873 ATCAAAGCCATTACAAACACTTTTGCAGTAAACCTCTACAACTCTTACCACTCCATTTG 932
QY 895 GCAGAGCTCGGATCTGCCACCAAAATGATCAAACTTTACCTTTGGGAATGACATGCC 954
Db 933 GCAGAGCTCGGATCTGCCACCAAAATGATCAAACTTTACCTTTGGGAATGACATGCC 992
QY 955 AATAGATATGACACTGTTTCATCCC-TCTCACCACCTTTCTTTCAAAATATCCGTTCA 1013
Db 993 AATAGATATGACACTGTTTCATCCC-TCTCACCACCTTTCTTTCAAAATATCCGTTCA 1052
QY 1014 CTACCTCATATGCAATTTCTATCTCTGAAAAGAAAGACGACCAATTAAGGGGGAATTCGA 1073
Db 1053 CTACCTCATATGCAATTTCTATCTCTGAAAAGAAAGACGACCAATTAAGGGGGAATTCGA 1110
QY 1074 GTTACCTGATGAGTTACAGGTGGGCGCTCTTTCATCCCAAGATTTCTCAAGCAT 1130
Db 1111 RKTACCGGATGGAGTACAAAGKGGC-TKCCCTCTVCATCCC-ARATWCTCAAGCAT 1165

```

RESULT 5

BX370812

LOCUS

DEFINITION

Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED

ACCESSION

BX370812

VERSION

BX370812.2

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 864)

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

On May 8, 2003 this sequence version replaced gi:30445558.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

4473.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0BAG062D02_CS05868_l&c=4473.r

FEATURES

source

1..864

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DL012YM24"

/cell_type="B CELLS (RAMOS CELL LINE)"

/cell_lines="RAMOS CELL LINE"

/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT

25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match

Best Local Similarity 52.9%; Score 793.6; DB 5; Length 864;

Matches 843; Conservative 0; Mismatches 9; Indels 4; Gaps 4;

```

QY 505 CCTTAATCTCATGAGGCAAGTCTCTGTTGGGTTAAAGAAAAGACCAAGTGGCA 564
Db 13 CCTTAATCTCATGCA-GCAAGTCTCTGTTGGGTTAAAGAAAAGC-CCAGTGGCA 70
QY 565 ACATTGTGATGCGCGTCTACCCGAAAGGATGTCGGAGACACATGTTGGTGCAC 624
Db 71 ACATTGTGATGCGCGTCTACCCG-AAGGATGTCGGAGACACATGTTGGTGCAC 129
QY 625 TGGAAAGAAAGACTTCTCTGTCAGTATTGTGCAGAGATTTGGGCGAAGATCACCT 684
Db 130 TGGAAAGAAAGACTTCTCTGTCAGTATTGTGCAGAGATTTGGGCGAAGATCACCT 189
QY 685 GATCGACATATGAAGAAGAGTCACAATCAAGAGCTTCTGAAGTCAAAACAGACCACT 744
Db 190 GATCGACATATGAAGAAGAGTCACAATCAGGAGCTTCTGAAGTCAAAACAGACCACT 249
QY 745 GGATTTCTTGACCAATTTACCTGCAATGTGTGTCCTTATAAAGAGAGAGCTCTTCC 804
Db 250 GGATTTCTTGACCAATTTACCTGCAATGTGTGTCCTTATAAAGAGAGAGCTCTTCC 309
QY 805 GGTGATGTCCTTACCTTCCAGTGAACCTGTATCAAGCCATTCAAAAACACTTTGAGTT 864
Db 310 GGTGATGTCCTTACCTTCCAGTGAACCTGTATCAAGCCATTCAAAAACACTTTGAGTT 369
QY 865 AAACCTCTACACACTTCCATTTTCACTTCCAGTCAAGAGCTCGGATCTGCCCAACATGAT 924
Db 370 AAACCTCTACACACTTCCATTTTCACTTCCAGTCAAGAGCTCGGATCTGCCCAACATGAT 429
QY 925 CACAACTTTACCTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCATCCCTCTCA 984
Db 430 CACAACTTTACCTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCATCCCTCTCA 489
QY 985 CCACCTTTCTTCAATATCGGTTTCACTTCTATGCAATTTCTATTCCTGAAAA 1044

```

490 CCACTTTCTTCAATATCCGTTTCAGTTCTACCTCATATGCAATTTCTATTCTGAAAA 549
1045 AGAACAGCCATTAAAGGGGAAATTTGAGAGTTACCTGATGAGTTACAGGTGCGTGCC 1104
550 AGAACAGCCATTAAAGGGGAAATTTGAGAGTTACCTGATGAGTTACAGGTGCGTGCC 609
1105 CTCTTATCCCAAGTTCTCAAGCATGTCATCATCTAAGCTAGGTTTGGATCCTCAGAT 1164
610 CTCTTATCCCAAGTTCTCAAGCATGTCATCATCTAAGCTAGGTTTGGATCCTCAGAT 669
1165 TGGTCCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1224
670 TGGTCCCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
1225 TGACCCCTTAAACACACACAGCATTTGGATTTTCTCAGTTGTTTAAATTTACCTTTAAA 1284
730 TGACCCCTTAAACACACAGCATTTGGATTTTCTCAGTTGTTTAAATTTACCTTTAAA 789
1285 TGGTCCCTTAAATTCCTTATCAGTGGGAGCCTTGAATGAGCTATTCCAGGAAGA 1344
790 TGGTCCCTTATATCTTATCTATCATGATGGGAGCCTTGAATGAGCTATTCCCA-GAGA 848
1345 AGCAGATCTCTGTT 1360
849 AGCAGATCTCTGTT 864

RESULT 6
AL036879 736 bp mRNA linear EST 04-SEP-2003
LOCUS DKFp564P1863_r1 564 (synonym: hfb2) Homo sapiens cDNA clone
DEFINITION DKFp564P1863 5', mRNA sequence.
ACCESSION AL036879
VERSION AL036879.3 GI:5927934
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J., and
Wiemann, S.
EST (Duesterhoeft, et al.)
Unpublished (1999)
On Jul 7, 1999 this sequence version replaced gi:5866332.
Contact: MIPS
MIPS
Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No si sequence available.
This clone (DKFp564P1863) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFp564P1863"
/tissue.type="brain"
/dev.stage="fetal"
/lab_host="X1-2blue"
/clone.lib="564 (synonym: hfb2)"
/notes="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

FEATURES
source
Query Match 44.1%; Score 662.2; DB 1; Length 736;
Best Local Similarity 98.1%; Pred. No. 3.9e-183;
Matches 723; Conservative 0; Mismatches 8; Indels 6; Gaps 5;

280 GTGTAAATTTGTGAGAAAAATGTTTACCGGAAAGATCATCTGAAGAAATCACCTCCATPAC 339
2 GGGTAAATTTATTTGTGAGAAAAATGTTTACCGGAAAGATCATCTGAAGAAATCACCTCCATPAC 61
340 ACAGACCCCTAAACAAAGAGAGCTTTAAAGTGGGAAAGATTTGGGCAAGAACTCAATACCAA 399
62 ACAGACCCCTAAACAAAGAGAGCTTTAAAGTGGGAAAGATTTGGGCAAGAACTCAATACCAA 121
400 GCTTGGATTTAAACGTCATCTTGGGCTTGCATGCGCAACAAAGTGGTGGTCACTCACTGTA 459
122 GCTTGGATTTAAACGTCATCTTGGGCTTGCATGCGCAACAAAGTGGTGGTCACTCACTGTA 181
460 GGTATGTTTGAACACTTTTGAAGCAGCGGAGTCTTCTGGAGCACCTTTAAATCTCATGC 519
182 GGTATGTTTGAACACTTTTGAAGCAGCGGAGTCTTCTGGAGCACCTTTAAATCTCATGC 241
520 AGGCAAGTCGTCTGTTGGGTTTAAAGAAAAAAGCACCAGTGCAGCAATTTGTGATCGCG 579
242 AGGCAAGTCGTCTGTTGGGTTTAAAGAAAAAAGCACCAGTGCAGCAATTTGTGATCGCG 301
580 GTTCTACACCCGAAAGAGTGTCCGGAGACACATGTTGTGGTGCACCTGGGAAGAAAGACTT 639
302 GTTCTACACCCGAAAGAGTGTCCGGAGACACATGTTGTGGTGCACCTGGGAAGAAAGACTT 361
640 CCTCTGTGATGTTGTGACACAGATTTGGGCGAAAGAGTCACTGACCTCGACATATGAA 699
362 CCTCTGTGATGTTGTGACACAGATTTGGGCGAAAGAGTCACTGACCTCGACATATGAA 421
700 GAAGAGTCAATCAAGAGCTTCTGAAGTCAAAACAGAAACAGAGTGGTTCCTTTCACCC 759
422 GAAGAGTCAATCAAGAGCTTCTGAAGTCAAAACAGAAACAGAGTGGTTCCTTTCACCC 481
760 ATTTACCTGCAATGTTCTGTGCTATATAAGAGCAGAGTCTCTCCGGTGAATGCTTACCC 819
482 ATTTACCTGCAATGTTCTGTGCTATATAAGAGCAGAGTCTCTCCGGTGAATGCTTACCC 541
820 TTCCAGTG-AACTGTTTATCAAGCCATTCAAAACACTTTGCAAGTTAAACCTCTTACAA 878
542 TTCCAGTGAAACTGTTTATCAAGCCATTCAAAACACTTTGCAAGTTAAACCTCTTACAA 601
879 CTCATTTCAGTCCATGC-AGAGTCCGGATCTGCCACC-AAATGATCAACATTTACCC 936
602 CTCATTTCAGTCCATGCAGAGCTCGGATCTGCCACC-CAAAATGATCAACATTTACCC 661
937 TTTGGGAATGACATGCCCAATAGATAT-GGACACTGTTTCCTCTCACCACCTTTCTT 995
562 TTTGGGAATGACATGCCCAATAGATATGGGACACTGTTCAATCTCTCAACACTTCTT 721
996 TCAATATCCGTTTCAGT 1012
722 --CAATATCCGTTTCAGT 736

RESULT 7
CA411972/c 734 bp mRNA linear EST 07-NOV-2002
LOCUS UI-H-EZ0-bal-m-01-0-UI.s1 NCI-CGAP_Chi Homo sapiens cDNA clone
DEFINITION UI-H-EZ0-bal-m-01-0-UI 3', mRNA sequence.
ACCESSION CA411972
VERSION CA411972.1 GI:24774623
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 734)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of Orthopaedics
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
 The following repetitive elements were found in this cDNA sequence: 1-22, >AT-rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Location/Qualifiers

1..734

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-H-EZO-bal-m-01-0-U"

/tissue_type="Chondrosarcoma Grade II"

/dev_stage="Adult"

/lab_host="PH10B (Life Technologies)"

/clone_lib="NCI CGAP Ch1"

/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP Ch1 is a cDNA library containing the following tissue(s): Chondrosarcoma Grade II. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is TGATCAGCT.

TAG TISSUE=grade-2-chondrosarcoma

TAG LIB=UI-H-EZO

TAG_SEQ=ATCTAATATG"

ORIGIN

Query Match 36.2%; Score 543.4; DB 6; Length 734;
 Best Local Similarity 99.5%; Pred. No. 3.7e-148;
 Matches 555; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 945 TGACATGCCCAATAGATGACACTGTTTCATCCCTCTCACCACCTTTCTTCAATATC 1004
 DB 734 TGACATGCCCAATAGATGACACTGTTTCATCCCTCTCACCACCTTTCTTCAATATC 675
 QY 1005 CGTTCAGTTCTACCTCATATGCATTTCTATTCTGAAAGAACAGCCATTAAAGGG-G 1063
 DB 674 CGTTCAGTTCTACCTCATATGCATTTCTATTCTGAAAGAACAGCCATTAAAGGGNN 615
 QY 1064 GAAATTGAGAGTTACCTGATGGAGTTACAGGTGGGTGCGCTTTCATCCCAAGATTCT 1123
 DB 614 GAAATTGAGAGTTACCTGATGGAGTTACAGGTGGGTGCGCTTTCATCCCAAGATTCT 555
 QY 1124 CAAGCATCTCATCTAAGCTAGGTTGAGTTCCTCAGATTGGGTCCCTAGATGATGGT 1183
 DB 554 CAAGCATCTCATCTAAGCTAGGTTGAGTTCCTCAGATTGGGTCCCTAGATGATGGT 495
 QY 1184 CGAGGAGACTCTCCCTATCCAAAGCTCTACTCTCAGTGACGCCCTTAAACACCA 1243
 DB 494 CGAGGAGACTCTCCCTATCCAAAGCTCTACTCTCAGTGACGCCCTTAAACACCA 435
 QY 1244 GAATTGGATTTTCTCAGTTGTTTAAATTTCAATACCTTTAAATGGTCCCTTAAATCT 1303
 DB 434 GAATTGGATTTTCTCAGTTGTTTAAATTTCAATACCTTTAAATGGTCCCTTAAATCT 375
 QY 1304 CTATCAGTGGGAGCCTTGGATAGCTATTCCAGGAGAGAGCATTCTTCTGTTTCC 1363
 DB 374 CTATCAGTGGGAGCCTTGGATAGCTATTCCAGGAGAGAGCATTCTTCTGTTTCC 315
 QY 1364 CAGTCTCCACACAAACAGAGATCTTCAGGATCTTCGAAACATATAGGGCTTGGGTCT 1423

Db 314 CAGTCCCCCAACAAACACAGGATCTTCAGGATCTTCGAAACACTATAGGCTTGGGTCT 255
 QY 1424 CTGCATCTCACTGTCAGCAGCTTTTCCACAGCAGTTTAAAGCAAGTACACCTCCACGT 1483
 Db 254 CTGCATCTCACTGTCAGCAGCTTTTCCACAGCAGTTTAAAGCAAGTACACCTCCACGT 195
 QY 1484 TTCCATCAAGCTTTTTCAG 1501
 Db 194 TTCCATCAAGCTTTTTCAG 177

RESULT 8

B94684/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: CIT-HSP-2170018.TR

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: M13-21;

Class: BAC ends.

Location/Qualifiers

1..492

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="GDB:7103260"

/db_xref="taxon:9606"

/clone="2170018"

/sex="Male"

/cell_type="Sperm"

/clone_lib="CIT-HSP"

/note="Vector: pBelobAC11; Site 1: HindIII; Site 2: HindIII"

Query Match 28.9%; Score 434; DB 8; Length 492;

Best Local Similarity 98.9%; Pred. No. 5.3e-116;

Matches 437; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1055 TTAAAGGGGAAATTGAGAGTTTACCTGATGGAGTTTACAGGTGGGTGCGCTTTCATCC 1114

Db 492 TTAAAGGGGAAATTGAGAGTTTACCTGATGGAGTTTACAGGTGGGTGCGCTTTCATCA 433

QY 1115 CAAGATTTCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCCTCAGATTGGTCCCTTA 1174

Db 432 CAAGATTTCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCCTCAGATTGGTCCCTTA 373

QY 1175 GATGATGTGTCAGGAGACCTCTCCCTATCCAAAAGCTCTATCTCCATCAGTACCCCTTA 1234

Db 372 GATGATGTGTCAGGAGACCTCTCCCTATCCAAAAGCTCTATCTCCATCAGTACCCCTTA 313

AUTHORS	Mahairas,G.G., Wallace,J.C.,Smith,K., Swartzell,S., Holzman,I., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
PUBMED	10449764
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Sequence tagged Connector Plate: 2170 row: 0 column: 18 Class: BAC ends High quality sequence stop: 365.

FEATURES	source
	Location/Qualifiers
1. .365	
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/clone="plate=2170 Col=18 Row=0"	
/sex="male"	
/clone_lib="CIT Approved Human Genomic Sperm Library D"	
/note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in E-Coli DH10B"	

	Query Match	23.5%;	Score 352.2;	DB 8;	Length 365;
	Best Local Similarity	97.8%;	Pred. No. 5.9e-92;		
	Matches 357; Conservative	0;	Mismatches 8;	Indels 0;	Gaps 0;
QY	1093	AGGTGGCGTGCCCTTCTTCATCCCAGAATTCTCAAGCATCGTCAATCATCTAAGCTTAGGTT	1152		
Dd	365	AGGTGGCGTGCCCTTAITTCATCCCCAAGATTCTCAAGCATCGTCAATCATATAAGCTTAGGTT	306		
QY	1153	GGATCCTCAGATTGGTGCCTTACATGATGTGCAGGAGACCTCTCCCTATCCAAAAGTCT	1212		

[illegible]

QY	1453	CAGTT 1457
DG	5	CAGTT 1
RESULT 14		
BU260281		
LOCUS	BU260281	804 bp mRNA linear EST 26-NOV-2002
DEFINITION	603501940F1 CSEQCHN51 Gallus gallus CDNA clone ChEST420p10 5', mRNA sequence.	
ACCESSION	BU260281	
VERSION	BU260281.1 GI:25527548	
KEYWORDS	EST.	
SOURCE	Gallus gallus (chicken)	

ORGANISM	REFERENCE AUTHORS	TITLE JOURNAL MEDLINE PUBMED	COMMENT
Gallus gallus	Boardman, P.E.,		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Fong, W.T., Fickel, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.		
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;	A Comprehensive Collection of Chicken CDNAs		
Phasianinae; Gallus.	Curr. Biol. 12 (22), 1965-1969 (2002)		
1 (Bases 1 to 804)			
	Contact: Simon Hubbard		
	Department of Biomolecular Sciences		
	University of Manchester Institute of Science and Technology		
	(UMIST)		
	PO Box 88, Manchester, M60 1QD, UK		
	Tel: 01612008930		
	Fax: 01612360409		
	Email: Simon.Hubbard@umist.ac.uk.		

```

Location/Qualifiers
1. 804
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="ChEST420p10"
/dev_stage="22"
/lab_host="DH108"
/clone_lib="CSFQCHEN51"
/notes="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN	Query Match	23.1%	Score 347:	DB 5:	Length 804:
	Best Local Similarity	68.9%	Pred. No. 2.6e-30:		
	Matches	505;	Conservative	0;	Mismatches 225; Indels 3; Gaps 2;
QY	75	GTGGTGAACCAACCAAGAAAAACTTTCTCTGCCAACTGTGTGACAAAGCCCTTTAACA	134		
Db	20	GTGAAGAGACTGAAGTCAGGGAATATGCCAATGTGAATATCAGAGACCTCTCTTCACTA	79		
QY	135	GTGTTTGAGAAATTAAGGTTTCACCTCTACTCTCACAGGAGAGAGGCCCTACAAAGTGCA	194		
Db	80	ATGTGCAACAGCTGAGAAGCTCACACGGCTTCTCTACTGTAGACAGAAACCGTACACTGCC	139		
QY	195	TTCAACAAGACTGTCACCAAGGCCCTTGTCTCTAAGTACAAATTACAAGGCACATGGCTA	254		
Db	140	CTCAGCTGCACCTGTGCGAAAGCCTTTGCTTTCTAAGTACAGCTGTATAGGCATATGCCCA	199		
QY	255	CTCATTTCTCCTCAGAAAAACCCACAAGTGTAAATTTGTGAGAAATGTTTCACCGGAAG	314		
Db	200	CTCAGCTCTCTCAGAAGCCTCATCAGTGTATGTACTGTGAGAAGATGTTTCACCGGAAG	259		
QY	315	ATCATCTGAAGAATCACTCTCCATACACAGGACCTTACAAAGACAGCGTTTAAGTGCGAAG	374		
Db	260	ATCACTTTCGCAACCACTTCGAGCCAGATCCCAACAGGAGGCCCTTCACCTGTCTGT	319		
QY	375	AATGTGGCAAGACTACATACCAAGCTTGGATTTAAACAGTCACATGTGGCCCTTGCATGCCG	434		
Db	320	AGTGTGCAAGAACTACAAACAAGCTCGGTTTCAAGCGCATCATCTGGCCATGCACGCTG	379		

